Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Matúš Čellár

Detekce změn v lineárních modelech a bootstrap

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Zuzana Prášková, CSc.
Studijní program: Matematika
Studijní obor: Pravděpodobnost, matematická statistika a ekonometrie

Praha 2016

Za nevyčerpatelnou trpělivost bych chtěl poděkovat paní doc. RNDr. Zuzaně Práškové, CSc. a za neutuchající motivaci své rodině a přátelům.

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V .... dne .... Podpis autora

Název práce: Detekce změn v lineárních modelech a bootstrap

Autor: Matúš Čellár

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Zuzana Prášková, CSc., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tématem práce jsou změny parametrů v lineárních modelech a metody jejich detekce. Práce začíná představením jejich dvou základních typů a bootstrapových procedur, které byly navrženy speciálně pro použití se závislými daty. V další kapitole se zaměříme na lokační model - nejjednodušší příklad lineárního modelu s uvažovanou změnou parametrů. Na tomto modelu si ukážeme způsob odhadování asymptotického rozptylu a implementaci zvolených bootstrapových procedur. V poslední kapitole si ukážeme jak použité metody upravit abychom je mohli použít v případě obecného lineárního modelu se změnou v parametrech. Na simulačních studiích pro oba uvažované modely porovnáme účinnost testů na změnu parametrů založených na asymptotických a bootstrapových kritických hodnotách. Taky prozkoumáme účinnost odhadu asymptotického rozptylu v situacích, když změna v parametrech nastane a když nenastane.

Klíčová slova: lineární model, detekce změn, simulační studie, bootstrap, závislá pozorování

Title: Change point detection in linear models and bootstrap

Author: Matúš Čellár

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Zuzana Prášková, CSc., Department of Probability and Mathematical Statistics

Abstract: This thesis discusses the changes in parameters of linear models and methods of their detection. It begins with a short introduction of the two basic types of change point detection procedures and bootstrap algorithms developed specifically to deal with dependent data. In the following chapter we focus on the location model - the simplest example of a linear model with a change in parameters. On this model we will illustrate a way of long-run variance estimation and implementation of selected bootstrap procedures. In the last chapter we show how to extend the applied methods to linear models with a change in parameters. We will compare the performance of change point tests based on asymptotic and bootstrap critical values through simulation studies in both our considered methods. The performance of selected long-run variance estimator will also be examined both for situations when the change in parameters occurs and when it does not.

Keywords: linear model, change point detection, simulation study, bootstrap, dependent observations

Contents

Introduction

Linear regression models are an example of results of mathematical statistics that achieved widespread application. They are utilized across a broad range of disciplines - from medicine, pharmaceutics and epidemiology through financial analysis and media research to hydrology and meteorology. Most of their applications fall into one of two broad categories - prediction and quantification of dependence.

When working with linear models one has to account for the possible changes in model parameters over the observation period. Change point detection procedures have been developed in order to help with this specific problem. There are two main types of these methods - off-line and on-line. Off-line procedures assume that all of the observations are known prior to the analysis. They correspond to the classic situation when we have a set of observations and we want to know whether regression parameters changed at any unknown historical time point. They are used to solve mainly hydrological, meteorological and some econometric problems. Recently an increasing number of data sets are collected automatically in such a way that the observations arrive in a steady pace (CAPM models, monitoring of intensive care patients, ...). With each new observation the question arises whether the model is still capable of explaining the data. This leads to sequential statistical analysis and so-called on-line methods. The choice between off-line and on-line methods depends on the specific problem and the question we are interested in.

Experience shows that the approximation of the distribution of the test statistics by its asymptotic limit is not sufficient - the convergence is too slow and they do not hold the overall significance level, especially when we do not have enough observations. Bootstrap modifications help to tackle these issues. A lot of work has already been done concerning these topics. For example in Antoch and Hušková (2003) bootstrapping modification in the off-line setting is discussed and in Hušková and Kirch (2012) the specifics of on-line setting are shown. The considered regression models in these articles have one common feature - they deal only with independent identically distributed model errors. This condition was relaxed in Antoch et al. (1997) where the impact on the asymptotic distribution of the test statistic was examined. Block bootstrap methods in change point detection problem with dependent residuals were discussed for example in Kirch (2007). In this thesis we aim to explore how we can utilize the bootstrap principle in this problem even further and compare the performance of these methods to the asymptotic approach. It will prompt the use of more complex bootstrapping procedures developed for dependent data - circular block and stationary block bootstrap, and dependent wild (or so-called jackknife) bootstrap.

The first chapter of this thesis describes the off-line change point detection problem based on Antoch and Hušková (2003). The second one is dedicated to the specifics of the on-line setting as discussed in Hušková and Kirch (2012). The following chapter provides a brief review of the bootstrap methods applicable to data displaying a dependency structure as provided in Kreiss and Paparoditis (2011). The fourth chapter focuses on the reader's introduction into the location model with dependent errors - a basic change point detection problem with non-trivial dependence structure, which was also analyzed in Antoch et al. (1997). In order to estimate long-run variance of our residuals flat-top kernels discussed in Politis (2001) were utilized. From the various options of kernel bandwidth selection we decided to use empirical rule-of-thumb introduced in Politis (2003) based on its simplicity and intuitive interpretation. The simulation study in the following chapter aims on the comparison of the performance of tests based on asymptotic and bootstrap-derived critical values. In the last chapter the methods are extended to linear models and a simulation study is performed on a regression line model with a possible change in intercept and slope.

Chapter 1

Off-line procedure with independent model errors

This method assumes that we know all of the observations prior to the change point detection analysis. It then estimates model parameters from the available data using e.g. the least squares method and tries to find the point, where the change occurred. One of the typical forms of the test statistic is based on the maximum of properly standardized partial sums of the estimated residuals and the point q where it is achieved is the prime suspect for our change point. If the value of the statistic manages to overstep a certain threshold based on the significance level, we declare that there was a significant change in model parameters at point q. The asymptotic results are obtained by letting the sample size, which will be denoted by n, go to infinity. First let us consider the model introduced in Antoch and Hušková (2003).

1.1 Model definition

The typical regression model used in off-line procedures has the form

$$y(t) = \boldsymbol{x}(t)^\top \boldsymbol{\beta}_n + \boldsymbol{x}(t)^\top \boldsymbol{\delta}_n \cdot I\{t > q_n\} + e(t), \quad t = 1, \ldots, n, \tag{1.1}$$

where $q_n (\leq n)$, $\boldsymbol{\beta}_n = (\beta_{1,n}, \ldots, \beta_{p,n})^\top$ and $\boldsymbol{\delta}_n = (\delta_{1,n}, \ldots, \delta_{p,n})^\top \neq 0$ are unknown parameters, $\boldsymbol{x}(t) = (\boldsymbol{x}_1(t), \ldots, \boldsymbol{x}_p(t))^\top, x_1(t) = 1, t = 1, \ldots, n$ are known design points and $e(t), \ldots, e(n)$ are *independent identically distributed* errors with zero mean, nonzero variance σ^2 and finite moment $\mathsf{E}\,|e(t)|^{2+\Delta_1}$ for some $\Delta_1 > 0$. Function $I\{A\}$ denotes the indicator of set A. The parameter q_n is called the *change point*.

1.2 Null hypothesis and test statistic

We are interested in the test of hypothesis

$$H_0 : q_n = n \text{ against } H_1 : q_n < n.$$

Clearly the null hypothesis corresponds to the situation where the regression parameters did not change in the observed sample. The test procedure not only advises us whether to reject the null hypothesis, but also provides us with an estimate of the change point q_n. It is based on the partial sums

$$\boldsymbol{S}_{kn} = \sum_{t=1}^{k} \boldsymbol{x}(t)\left(y(t) - \boldsymbol{x}(t)^\top \hat{\beta}_n\right) \qquad ,k = 1, \dots, n,$$

$$S^*_{kn} = \sum_{t=1}^{k} \left(y(t) - \boldsymbol{x}(t)^\top \hat{\beta}_n\right) \qquad ,k = 1, \dots, n,$$

where

$$\hat{\beta}_n = \boldsymbol{C}_{nn}^{-1} \sum_{t=1}^{n} \boldsymbol{x}(t) y(t), \quad C_{kn} = \sum_{t=1}^{k} \boldsymbol{x}(t)\boldsymbol{x}(\boldsymbol{t})^\top, \quad k = 1, \dots, n,$$

$$C^0_{kn} = C_{kn} - C_{nn}, \quad k = 1, \dots, n.$$

We are in fact working with partial (weighted) sums of L_2-residuals

$$\hat{e}(t) = y(t) - \boldsymbol{x}(t)^\top \hat{\beta}_n, \quad t = 1, \dots, n.$$

The test statistic based on $\boldsymbol{S}_{kn}$ for testing H_0 against H_1 has the form

$$T_n = \max_{p<k<n-p} \{\hat{\sigma}_n^{-2} \boldsymbol{S}^\top_{\boldsymbol{kn}} \boldsymbol{C}^{\boldsymbol{-1}}_{\boldsymbol{kn}} \boldsymbol{C}_{\boldsymbol{nn}} \boldsymbol{C}^{\boldsymbol{0^{-1}}}_{\boldsymbol{kn}} \boldsymbol{S}_{kn}\},$$

where $\hat{\sigma}_n^2$ is an estimator of σ^2 with the property

$$\hat{\sigma}_n^2 - \sigma^2 = o_{\mathsf{P}}((\log\log n)^{-\frac{1}{2}}), n \to \infty$$

The test statistic T_n is related to the likelihood ratio test when the errors have $N(0, \sigma^2)$ distribution. The term $\hat{\boldsymbol{\beta}}_n$ in fact represents the least squares estimator of $\boldsymbol{\beta}$ in the model with $m = n$, i.e. the case with no change in the regression parameters. It is also important to point out that the model only allows a change in regression parameters, i.e. a change in $\boldsymbol{\beta}$, the variance parameter σ^2 is assumed to be constant. For its estimation Antoch and Hušková (2003) recommend

$$\hat{\sigma}_n^2 = \frac{1}{n-p} \left\{ \sum_{t=1}^{n} \hat{e}^2(t) - \max_{p<k<n-p} \boldsymbol{S}^\top_{kn} \boldsymbol{C}^{-1}_{kn} \boldsymbol{C}_{nn} \boldsymbol{C}^{\boldsymbol{0^{-1}}}_{\boldsymbol{kn}} \boldsymbol{S}_{\boldsymbol{kn}} \right\}. \tag{1.2}$$

The test statistic based on $S^*_{kn}, k = 1, \dots, n$, is of the form

$$T^*_n = \max_{1 \le k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \cdot \frac{|S^*_{kn}|}{\hat{\sigma}_n} \right\}. \tag{1.3}$$

The quite complex estimator of σ^2 can be replaced by a simpler one if computational simplicity is preferred

$$\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{e}^2(t).$$

Large values of the above specified test statistic indicate that the null hypothesis is violated. The critical regions are therefore of the form

$$T_n > c_n(\alpha) \text{ and } T_n^* > c_n^*(\alpha),$$

where $c_n(\alpha)$ and $c_n^*(\alpha)$ are critical values corresponding to the level α. These values can be approximated through the limit distribution of the respective test statistic under the null hypothesis.

1.3 Limit distribution under the null hypothesis

1.3.1 Assumptions

In this section we formulate the assumptions used in Antoch and Hušková (2003) in order to arrive to the limit distribution of our test statistics. Concerning *the design points* $\boldsymbol{x}(t), t = 1, \ldots, n$ we assume:

[A.1.] $x_1(t) = 1, \quad t = 1, \ldots, n$ and $\sum_{t=1}^{n} x_j(t) = 0, \quad j = 2, \ldots, p.$

[A.2.] There exists a positive definite $p \times p$ matrix $\boldsymbol{C}$ such that for any sequence $\{l_n\}$, $\lim_{n\to\infty} l_n = \infty$, $l_n \le n$ and

$$\left\| \frac{1}{l_n}(\boldsymbol{C}_{k+l_n n} - \boldsymbol{C}_{kn}) - \boldsymbol{C} \right\| = o((\log l_n)^{-1})$$

uniformly for $1 \le k \le n - l_n$, where $\|\cdot\|$ denotes the Euclidean norm.

[A.3.] For $n \to \infty$

$$\max_{1\le k\le n} \left\{ \frac{1}{k} \sum_{t=1}^{k} \|\boldsymbol{x}(t)\|^4 + \frac{1}{n-k} \sum_{t=k+1}^{n} \|\boldsymbol{x}(t)\|^4 \right\} = O(1).$$

For the distribution of the *error terms* $e(t)$ we assume the following:

[B.1.] $e(1), e(2), \ldots$ are *i.i.d. random variables* with zero mean, non-zero variance σ^2 and finite moment $\mathsf{E}\, |e(t)|^{2+\Delta_1}$ for some $\Delta_1 > 0$.

1.3.2 Asymptotic results

Under these assumptions it can be shown (the proof can be found in the third chapter of Csörgö and Horváth (1997)) that the limiting distributions have the following form

$$\lim_{n\to\infty} \mathsf{P}\left(a(\log n)\sqrt{T_n} \le y + b_p(\log n) \right) = \mathrm{e}^{-2\,\mathrm{e}^{-y}}, \quad y \in \mathbb{R}, \tag{1.4}$$

$$\lim_{n\to\infty} \mathsf{P}\left(a(\log n) T_n^* \le y + b_1(\log n) \right) = \mathrm{e}^{-2\,\mathrm{e}^{-y}}, \quad y \in \mathbb{R}, \tag{1.5}$$

where

$$a(t) = \sqrt{2\log t}, \quad b_p(t) = 2\log t + \frac{p}{2}\log\log t - \log\left(\Gamma\left(\frac{p}{2}\right)\right), \quad t > 1$$

and $\Gamma(\cdot)$ stands for Gamma function. This result allows us to construct asymptotic critical values and asymptotic p-values for the proposed tests. Even for more general statistics with a weight function w of the form

$$T_n(w) = \sup_{0<t<1} \left\{ \frac{\boldsymbol{S}_{\boldsymbol{kn}}^\top \boldsymbol{C}_{\boldsymbol{kn}}^{\boldsymbol{-1}} \boldsymbol{C}_{\boldsymbol{nn}} \boldsymbol{C}_{\boldsymbol{kn}}^{\boldsymbol{0}^{\boldsymbol{-1}}} \boldsymbol{S}_{kn}}{w(t)^2 \cdot \hat{\sigma}_n^2} \right\},$$

$$T_n^*(w) = \sup_{0<t<1} \left\{ \frac{\left| S^*_{\lfloor (n+1)t \rfloor n} \right|}{\sqrt{n} \cdot w(t) \cdot \hat{\sigma}_n} \right\},$$

similar results can be achieved. However, the weight function needs to satisfy a few reasonable assumptions:

[C.1.] $w : (0,1) \to (0,\infty)$ needs to be non-decreasing in a neighborhood of 0, non-increasing in a neighborhood of 1, $\inf\{w(t) : t \in \{\eta, 1-\eta\}\} > 0, \forall \eta \in (0, \frac{1}{2})$ and for some $c > 0$

$$\int_0^1 \frac{1}{s(1-s)} \exp - \frac{cw^2(s)}{s(1-s)} \,\mathrm{d}s < \infty.$$

If all of the assumptions [A.1.]-[A.3.], [B.1.] and [C.1.] are satisfied then, as $n \to \infty$

$$\sqrt{T_n(w)} \xrightarrow{\mathcal{D}} \sup_{0<t<1} \left\{ \frac{\sqrt{\sum_{i=1}^p B_i^2(t)}}{w(t)} \right\} \text{ and } T_n^*(w) \xrightarrow{\mathcal{D}} \sup_{0<t<1} \left\{ \frac{|B_1(t)|}{w(t)} \right\},$$

where $\{B_j(t) : t \in (0,1)\}_{j=1}^p$ are independent Brownian bridges. Proof can be found in the third chapter of Csörgö and Horváth (1997).

However the asymptotic critical values obtained from these approximations are not sufficient. The convergence to the limiting distributions tends to be slow and tests based on them have trouble with holding level. In Antoch and Hušková (2003) the assumption [B.1.] allowed the use of permutation principle, which corresponds to residual bootstrap, to improve these critical values making use of computer simulations.

Chapter 2

On-line procedure with independent model errors

The on-line method finds great use in situations where the observations arrive in a steady pace. It assumes that we have a historical data set with no change before the monitoring starts. The least squares method is used in this set to estimate the model parameters. Then, as the new observations appear, we are interested whether the model is still capable of explaining the arriving data. The test statistic is again based on properly standardized partial sums of the estimated residuals which are compared to a threshold. Once it is reached, an alarm is raised, the procedure is stopped and the model parameters need to be adjusted. In on-line procedure the threshold is updated with each new arriving observation. It needs to be done in such a way that the overall significance level is controlled. Let us assume that the monitoring will continue to infinity, if no alarm is raised, then we end up with so-called *open-end* on-line procedures. In many situations it is more realistic, that the monitoring will stop after a finite time horizon even if no change is detected. That is the case of *closed-end* on-line procedures. The asymptotic results are obtained by letting the size of the historic data set, denoted by m, go to infinity. The main source used for this chapter was Hušková and Kirch (2012), where the following model was introduced.

2.1 Model definition

Let us consider a linear model

$$y(i) = \boldsymbol{x}(i)^\top \boldsymbol{\beta}_{i,m} + e(i), i \geq 1.$$

Let $\boldsymbol{x}(i)$ be a $p \times 1$ random vector of regressors and $\boldsymbol{\beta}_{i,m}$ is a $p \times 1$ vector of parameters. We assume again that the error sequence is *independent identically distributed* and independent of the regressors, has zero mean, nonzero variance σ^2 and finite moments $\mathsf{E}\,|e(i)|^{2+\Delta_1}$ for some $\Delta_1 > 0$. Parameter estimation is based on the historical data set where we assume that no change in regression parameters occurred, i.e.

$$\boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_0, i = 1, \dots, m.$$

2.2 Null hypothesis and test statistic

We are interested in testing the null hypothesis

$$H_0 : \boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_0, \quad m+1 \leq i < m+1+T_m$$

against the alternative of a change in the regression coefficients

$$\begin{aligned} H_1 : \text{ there is a } q_m \geq 1 \text{ such that } \boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_0, \quad m < i < m + q_m \\ \text{and } \boldsymbol{\beta}_{i,m} = \boldsymbol{\beta}_m^0 \neq \boldsymbol{\beta}_0, \quad m + q_m < i < m+1+T_m. \end{aligned}$$

T_m is the observation horizon which can be finite (closed-end) or infinite (open-end), but has to converge to infinity with m. The values $\boldsymbol{\beta}_0, \boldsymbol{\beta}_m^0, q_m$ are not known.

The test introduced in Horváth et al. (2004) and discussed in Hušková and Kirch (2012) is based on

$$\Gamma(m,k,\gamma) = \frac{\sum_{m<i\leq m+k} \left(y(i) - \boldsymbol{x}(i)^\top \hat{\beta}_m \right)}{g(m,k,\gamma)},$$

where

$$g(m,k,\gamma) = m^{\frac{1}{2}} \left(1 + \frac{k}{m}\right) \left(\frac{k}{m+k}\right)^\gamma$$

for $0 \leq \gamma < \frac{1}{2}$ and

$$\hat{\boldsymbol{\beta}}_m = \boldsymbol{C}_{mm}^{-1} \sum_{j=1}^{m} \boldsymbol{x}(j) y(j),$$

where $\boldsymbol{C}_{mm}$ is the same as in the previous chapter. We can observe that $\hat{\boldsymbol{\beta}}_m$ is the least squares estimator of the vector of regression coefficients based on the historical data set $y(1), \ldots, y(m)$. The test statistic is then given by

$$\tau_m(\gamma) = \frac{1}{\hat{\sigma}_m} \sup_{1\leq k<T_m+1} |\Gamma(m,k,\gamma)|,$$

where $T_m \to \infty, \frac{T_m}{m} \to N > 0$, as $m \to \infty$ (closed-end procedure), or $T_m = \infty$ (open-end procedure) and $\hat{\sigma}_m^2$ is a consistent estimator of σ^2 only depending on the historical data set. Often

$$\hat{\sigma}_m^2 = \frac{1}{m-p} \sum_{i=1}^{m} \left(y(i) - \boldsymbol{x}(i)^\top \hat{\boldsymbol{\beta}}_m \right)^2.$$

The null hypothesis is rejected at the following stopping time

$$\tau(m) = \begin{cases} \inf\{1 \leq k < T_m + 1 : \frac{1}{\hat{\sigma}_m} |\Gamma(m,k,\gamma)| \geq c\}, \\ \infty, \text{ if } \frac{1}{\hat{\sigma}_m} |\Gamma(m,k,\gamma)| < c \text{ for all } 1 \leq k < T_m + 1, \end{cases} \tag{2.1}$$

c is chosen in such a way that the false alarm is controlled, i.e. under the null hypothesis

$$\lim_{m\to\infty} \mathsf{P}(\tau(m) < \infty) = \alpha \tag{2.2}$$

needs to hold for our desired $0 < \alpha < 1$ significance level. We also require that under the alternative H_1

$$\lim_{m\to\infty} \mathsf{P}(\tau(m) < \infty) = 1, \tag{2.3}$$

i.e. we require the test to have asymptotic power one.

2.3 Limit distribution under the null hypothesis

2.3.1 Assumptions

In this section we formulate the assumptions used in Hušková and Kirch (2012) in order to arrive to the limit distribution of our test statistics. Concerning the covariates we assume the following:

[α .1.] For the sequence $\{\boldsymbol{x}(i) = (x_1(i), \ldots, x_p(i))^\top 1 \leq i < \infty\}$ there exists a positive definite matrix $\boldsymbol{C}$ and a constant $0 < \rho \leq \frac{1}{2}$ such that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}(i)\boldsymbol{x}(i)^\top - \boldsymbol{C} \right\|_\infty = O(n^{-\rho}) \text{ a.s.},$$

where $\| \cdot \|_\infty$ denotes the maximum norm of matrices.

Concerning the random errors we need similar assumptions as we had in the first chapter.

[β .1.] $\{e(i) : 1 \leq i < \infty\}$ are *i.i.d. random variables* with zero mean, nonzero variance σ^2 and finite moment $\mathsf{E}\, |e(i)|^{2+\Delta_1}$ for some $\Delta_1 > 0$.

[β .2.] The sequences $\{e(i) : 1 \leq i < \infty\}$ and $\{\boldsymbol{x}(i) : 1 \leq i < \infty\}$ are independent.

We also need the parameter γ to satisfy

[γ .1.] $0 \leq \gamma < \rho$, where ρ is the constant from assumption [α.1.].

And additionally we will also need an assumption for the observation horizon T_m in the case of the closed-end procedure

[δ .1.] If $T_m < \infty$, then $lim_{n\to\infty} \frac{T_m}{m} = N$ for some $0 < N < \infty$ or $lim_{n\to\infty} \frac{T_m}{m} = \infty$.

2.3.2 Asymptotic results

For the open-end procedure under the null hypothesis and under [α.1.], [β.1.], [β.2.] and [γ.1.] it has been shown in Horváth et al. (2004) that as $m \to \infty$

$$\lim_{m\to\infty} \mathsf{P}\left(\sup_{1\leq k<\infty} \frac{|\Gamma(m,k,\gamma)|}{\hat{\sigma}_m} \leq y \right) = \mathsf{P}\left(\sup_{0\leq t\leq 1} \frac{|W(t)|}{t^\gamma} \leq y \right), \quad \forall y \in \mathbb{R}, \tag{2.4}$$

where $\{W(t) : 0 \leq t < \infty\}$ denotes a Wiener process. It is worth noting that the explicit form of the limit distribution is known only for the case when $\gamma = 0$ and has to be simulated otherwise. Using quantiles of the limit distribution as critical values c ensures that (2.2) holds, i.e. the corresponding open-end test asymptotically controls the overall false-rejection rate.

If we also assume [δ.1.] then a slight modification of the proof of the assertion above shows that as $m \to \infty$

$$\mathsf{P}\left(\sup_{1\leq k<T_m+1} \frac{|\Gamma(m,k,\gamma)|}{\hat{\sigma}_m} \leq y \right) =$$
$$= \mathsf{P}\left(\sup_{1\leq k<T_m+1} \frac{\left|W_1(\frac{k}{m}) - \frac{k}{m}W_2(1)\right|}{(1+\frac{k}{m})(\frac{k}{k+m})^\gamma} \leq y \right) + o_{\mathsf{P}}(1), \quad \forall y \in \mathbb{R}, \tag{2.5}$$

where $\{W_1(t) : 0 \leq t < \infty\}, \{W_2(t) : 0 \leq t < \infty\}$ are two independent Wiener processes. Also it can be shown that the distribution of the right hand side of (2.4) converges to the same limit as given in (2.5) if $\frac{T_m}{m} \to \infty$. On the other hand, if $\frac{T_m}{m} \to N$ then the right hand side of (2.4) converges to the distribution of

$$\sup_{0 \leq t \leq \frac{N}{N+1}} \frac{|W(t)|}{t^\gamma}.$$

Again the quantiles of this distribution have to be simulated, but they are very close to the quantiles given by the limit distribution in (2.4) for $N \geq 10$. Using the quantiles of the limit distribution in (2.5) as the critical values c in (2.1) ensures that the overall asymptotic false-alarm rate is controlled.

Concerning the asymptotic power of the tests several results have been proven. For the case when $\boldsymbol{\beta}_m^0 = \boldsymbol{\beta}^0$ and $\boldsymbol{c}_1^\top(\boldsymbol{\beta}^0 - \beta_0) \neq 0$, where $\boldsymbol{c}_1$ stands for the first column of $\boldsymbol{C}$, Horváth et al. (2004) proved that

$$\frac{1}{\hat{\sigma}_m} \sup_{1 \leq k < \infty} |\Gamma(m, k, \gamma)| \xrightarrow{\mathsf{P}} \infty.$$

Hence the open-end monitoring procedure has asymptotic power one. Assuming $\limsup_{m \to \infty} \frac{q_m}{m} < N$, where $N = \lim_{m \to \infty} \frac{T_m}{m}$, the result can be extended to the closed-end monitoring procedure and both of these results can even be extended to include a more general form of alternatives with less restrictions.

In Hušková and Kirch (2012) the assumption $[\beta.1.]$ allowed the use of residual and pair bootstrap in order to find bootstrap critical values for the proposed tests and compare them to the critical values obtained from the limit distributions. The simulation study showed the strengths of bootstrap critical values, especially for small historical sample sizes. The aim of this thesis is to perform similar study on models which allow a dependence structure for the model errors. In order to perform bootstrap simulations on such data we will need to work with more sophisticated bootstrap methods developed specifically for dependent data.

Chapter 3

Bootstrap methods for dependent data

Bootstrap procedures focus mainly on approximating moments or distribution of a properly standardized estimator T_n (which depends on the observations $X_1, \ldots, X_n$ of a given time series) of a parameter ν. Typically a proper standardization of the estimator means that we find a sequence $\{c_n : n \in \mathbb{N}\}$ of non-negative real numbers such that the sequence of distributions $\mathcal{L}_n = \mathcal{L}\left(c_n(T_n - \nu)\right)$ converges to a non-degenerate limit. Computer simulations are used in order to generate a significant number of bootstrap datasets derived from our observations $X_1, \ldots, X_n$ and values of the estimator T_n on these generated pseudo-observations. These values are then used to empirically estimate the desired distribution $\mathcal{L}\left(c_n(T_n - \nu)\right)$. A number of bootstrap methods was specifically designed to work with data which display a dependence structure. The aim of this chapter is to introduce the most common ones. All of the presented procedures assume that we have observations $X_1, \ldots, X_n$ at hand stemming from a stationary process $\{X_t : t \in \mathbb{Z}\}$ with zero mean and finite second-order moments. The autocovariance function $R : \mathbb{Z} \to \mathbb{R}$, where $R(k) = \mathsf{E}\left(X_t, X_{t+k}\right)$, needs to fulfill $R(0) > 0$ and $\sum\limits_{k \in \mathbb{Z}} |R(k)| < \infty$. A review of the procedures working with dependent data, which was the main source for this chapter, can be found in Kreiss and Paparoditis (2011).

3.1 Block bootstrap

The block bootstrap is a method for extending original bootstrap idea of drawing with replacement to dependent time series observations. The main idea of all block bootstrap procedures consists of dividing data into blocks of consecutive observations of length l, $(X_t, X_{t+1}, \ldots, X_{t+l-1})$ say, and sampling the blocks randomly from all possible blocks. One can consider overlapping or non-overlapping blocks and random or non-random length l. It is interesting to point out that when random block length is used, then the resulting bootstrap series itself is stationary.

3.1.1 Block bootstrap with fixed block length

The basic resampling algorithm (described in Kreiss and Paparoditis (2011)) used in the block bootstrap with non-random block length has two main steps:

1. Let $l \in \mathbb{N}, l \ll n, L = \left\lfloor \frac{n}{l} \right\rfloor$ and $k = n - L \cdot l$. Define discrete, independent random variables $t_1, t_2, \ldots, t_{L+1}$ taking values in the set $I_{n,l}$ where

   (a) $I_{n,l} = \{1, 2, \ldots, n - l + 1\}$ if overlapping blocks

   (b) $I_{n,l} = \{1, l + 1, 2l + 1, \ldots, (L - 1)l + 1\}$ if non-overlapping blocks

   are considered.

2. Lay the blocks $(X_{t_s}, X_{t_s+1}, \ldots, X_{t_s+l-1}), s = 1, \ldots, L + 1$ end to end in the order sampled together and discard the last $l - k$ observations to form a bootstrap pseudo-series $X_1^*, \ldots, X_n^*$.

The block bootstrap approximation of the distribution $\mathcal{L}_n = \mathcal{L}\left(c_n(T_n - \nu)\right)$ is then given by $\mathcal{L}_n^* = \mathcal{L}\left(c_n(T_n^* - \nu^*)\right)$, where $T_n^* = T_n(X_1^*, \ldots, X_n^*)$ and ν^* denotes some properly chosen centering parameter.

3.1.2 Stationary bootstrap

In the case of block bootstrap with random block length, namely the stationary bootstrap, Kreiss and Paparoditis (2011) adjusted the steps of the algorithm as follows:

1.* The lengths L_i of the blocks to be selected consist of i.i.d. random variables having a geometric distribution with parameter $p \in (0, 1)$.

2.* The first L_1 pseudo-observations of the bootstrap time series $X_1^*, \ldots, X_n^*$ consist of observations $X_{t_1}, \ldots, X_{t_1+L_1}$, the next L_2 bootstrap observations are the observations of the second sampled block of random length L_2 and so on. The bootstrap data generating process is stopped once n bootstrap observations have been generated.

Alternatively the resampling scheme for the stationary bootstrap can be described in the following simpler way:

Let X_1^* be randomly chosen from the underlying sample. Now, given that X_i^* is equal to the observation X_t of the underlying sample, with probability $1 - p$ the next bootstrap observation X_{i+1}^* is chosen to be equal to X_{t+1} and with probability p it is randomly chosen from the whole underlying sample $X_1, \ldots, X_n$.

The bootstrap time series generated in this way is indeed stationary and shares a nice Markovian property. For further details we refer to Politis and Romano (1994). The approximation of the desired distribution $\mathcal{L}_n$ is then done similarly as in the previous case.

3.2 Non-parametric residual bootstrap

The original bootstrap idea of drawing with replacement from i.i.d. random variables can be utilized also in case of dependent observations by applying the original bootstrap principle to residuals of an optimal predictor of the X_t. Consider a fixed $p \in \mathbb{N}$ and denote a non-parametric estimator of the conditional expectation $\mathsf{E}[X_t|X_{t-1},\dots,X_{t-p}]$ by $\hat{m}_n(X_{t-1},\dots,X_{t-p})$. This approach leads to residuals

$$\hat{e}_t = X_t - \hat{m}_n(X_{t-1},\dots,X_{t-p}), \quad t = p+1,\dots,n$$

and to a bootstrap time series

$$X_t^* = \hat{m}_n(X_{t-1},\dots,X_{t-p}) + e_t^*, \quad t = p,\dots,n.$$

One needs to decide on the way of choosing the initial bootstrap observations $X_1^*,\dots,X_p^*$. The bootstrap innovations $e_1^*,\dots,e_n^*$ are assumed to possess a uniform distribution on the set $\{\hat{e}_{p+1}^C,\dots,\hat{e}_n^C\}$ of centralized residuals

$$\hat{e}_t^C = \hat{e}_t - \frac{1}{n-p}\sum_{k=p+1}^{n}\hat{e}_k, \quad t = p+1,\dots,n.$$

3.3 The dependent wild bootstrap

The dependent wild bootstrap introduced in Shao (2010) is an extension of the traditional wild bootstrap introduced in Wu (1986) to the time series setting by allowing the auxiliary variables involved in the wild bootstrap to be dependent, so it is capable of mimicking the dependence structure in the original time series. It is superior to block-bootstrap methods when dealing with time series containing missing observations or unequally spaced data.

According to our assumptions introduced in the beginning of the chapter we have a stationary series $\{X_t, t \in \mathbb{Z}\}$ with finite variance σ^2. Let us denote $\mu = \mathsf{E}(X_t)$ the mean and $R(k)$ the autocovariance function. Given the observations $X_1,\dots,X_n$ from time series $\{X_t, t \in \mathbb{Z}\}$ Shao (2010) defines the dependent wild bootstrap pseudoseries as

$$X_t^* = \bar{X}_n + (X_t - \bar{X}_n)\cdot W_t, \quad t = 1,\dots,n, \tag{3.1}$$

where $\bar{X}_n = \frac{1}{n}\sum_{s=1}^{n} X_s$ is the sample mean and $\{W_t\}_{t=1}^n$ are random variables satisfying the following assumption.

[A.1.] The random variables $\{W_t\}_{t=1}^n$ are independent of our data, $\mathsf{E}(W_t) = 0$, $\mathrm{var}(W_t) = 1$ for $t = 1,\dots,n$. Assume that $\{W_t\}$ is a stationary process with autocovariance function $R(k) = cov(W_t, W_{t+k}) = a(\frac{k}{l})$, where $a(\cdot)$ is a kernel function and $l = l_n$ is a bandwidth parameter. Furthermore, assume that

$$K_a(x) = \int_{-\infty}^{\infty} a(u)\,\mathrm{e}^{-iux}\,\mathrm{d}u \geq 0, x \in \mathbb{R}. \tag{3.2}$$

The bandwidth parameter l plays a similar role as the block size in the block-based methods. Note that assumption (3.2) is satisfied by a few commonly used kernels like Bartlett, Parzen and quadratic spectral kernels. They can be found in Andrews (1991), where their performance in covariance matrix estimation is compared.

The term "dependent wild bootstrap" was chosen based on two considerations. On one hand, it is akin to the wild bootstrap, which was originally introduced to deal with independent and heteroscedastic errors in the regression problems. On the other hand, unlike the traditional wild bootstrap, the random variables $\{W_t\}_{i=1}^n$ here are dependent, which allows us to capture the dependence structure in the original sample, hence the dependent wild bootstrap.

Chapter 4

Location model with dependent residuals

For the purpose of illustrating the behavior and performance of bootstrap procedures in the change point detection problem, let us consider a special case of a model with a change in parameters and dependent residuals - location model with a change in the mean and residuals that form a linear process. This model was also considered in Antoch et al. (1997), which was focused mainly on answering the question how the allowed dependence structure affects the asymptotic distribution of the test statistic and evaluating how well do asymptotic critical values perform in the detection of change in the mean setting. This thesis aims on discussing an alternative to the asymptotic approach by making use of bootstrap procedures designed to work with dependent data. The upcoming chapter will focus on the reader's introduction into this specific problem and the implementation of these procedures.

4.1 Model definition

Let us consider the following model, which was also discussed in Antoch et al. (1997),

$$\begin{aligned} y(t) &= \mu + e(t), & t &= 1, \ldots, q_n, \\ &= \mu + \delta + e(t), & t &= q_n + 1, \ldots, n, \end{aligned} \tag{4.1}$$

where $\mu \in \mathbb{R}$, $\delta \neq 0$ and $1 \leq q_n \leq n$ are parameters. The error terms $\{e(t)\}_{t=1}^{\infty}$ are assumed to satisfy the following assumptions.

[b.1.] The error terms $\{e(t)\}_{t=1}^{\infty}$ behave as a linear process, i.e.

$$e(t) = \sum_{j=0}^{\infty} w_j \epsilon_{t-j}, \quad t = 1, 2, \ldots \tag{4.2}$$

where $\{\epsilon_s\}_{s=-\infty}^{\infty}$ are i.i.d. random variables with $\mathsf{E}\ \epsilon_s = 0$, $\operatorname{var} \epsilon_s = \sigma^2 > 0$ and $\mathsf{E}\ |\epsilon_s|^{2+\Delta} < \infty$ for some $\Delta > 0$ and the weights $\{w_j\}_{j=0}^{\infty}$ satisfy

$$\sum_{j=0}^{\infty} j\,|w_j| < \infty\ , \quad \sum_{j=0}^{\infty} w_j \neq 0. \tag{4.3}$$

4.2 Null hypothesis and the test statistic

The interest lies in testing the null hypothesis

$$H_0 : q_n = n \text{ against } H_1 : q_n < n,$$

i.e. whether there occurred a change in the mean of the series or not. The test statistic used in Antoch et al. (1997) was of the form

$$\tilde{T}_n = \max_{1 \le k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \, |S_{kn}| \right\},$$

where S_{kn} stands for k-th partial sum of the ordinary least squares residuals from the model with $q_n = n$, i.e. $S_{kn} = \sum_{t=1}^{k} (y(i) - \bar{y}_n)$. In the aforementioned article the following asymptotic properties were proven.

Theorem 1. *If H_0 holds and the assumptions* (4.1), (4.2) *and* (4.3) *are met, then*

$$\lim_{n \to \infty} P\left(a(\log n) \widetilde{T}_n \le \sigma^* (y + b_1(\log n)) \right) = \mathrm{e}^{-2\,\mathrm{e}^{-y}}, \tag{4.4}$$

where

$$a(y) = \sqrt{2 \log y} \textit{ and } b_p(y) = 2 \log y + \frac{p}{2} \log \log y - \log \left(\Gamma \left(\frac{p}{2} \right) \right), y > 1,$$

and

$$\sigma^{*2} = \lim_{n \to \infty} var \left(\sqrt{n} \cdot (\bar{y}_n - \mu) \right)$$

is the long-run variance.

Note. Under the null hypothesis and the mentioned assumptions (4.1), (4.2) and (4.3) the process $\{y(t)\}_{-\infty}^{\infty}$ is stationary and has spectral density function f of the form

$$f(\omega) = \frac{1}{2\pi} \cdot \sum_{k=-\infty}^{\infty} \mathrm{e}^{i\omega k} \, R(k), \quad \omega \in [-\pi, \pi].$$

Furthermore, under our assumptions the long-run variance satisfies

$$\sigma^{*2} = \lim_{n \to \infty} \mathrm{var} \left(\sqrt{n} \cdot (\bar{y}_n - \mu) \right) = 2\pi \cdot f(0).$$

In order to make use of this theorem for testing the null hypothesis based on asymptotic critical values, σ^* needs to be estimated. In Antoch et al. (1997) the following estimate is recommended

$$\hat{\sigma}_n^2(L) = \hat{R}(0) + 2 \sum_{k=1}^{\lfloor L \rfloor} \left(1 - \frac{k}{L} \right) \hat{R}(k), \quad L < n \tag{4.5}$$

where

$$\hat{R}(k) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n-k}(y(i)-\bar{y}_n)(y(i+k)-\bar{y}_n), & k = 0,\dots,n-1, \\ \hat{R}(-k), & k = -1,\dots,-n+1, \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

and $\bar{y}_n = \frac{1}{n}\sum_{j=1}^{n} y(j)$ denotes the mean of our observations.

Note. It should be pointed out that the estimates $\hat{R}(k)$ depend on n and should be denoted as $\hat{R}_n(k)$, but the index is omitted in order to keep the notation simpler.

Even a more robust version of the estimator which takes into account a possible change in model parameters is introduced in Antoch et al. (1997). It is important that the article also mentions that if the lag window L is chosen correctly, the asymptotic property (4.4) remains valid even if we replace σ^* with $\hat{\sigma}_n(L)$. More precisely if $\frac{L^2}{\log n} \to \infty$ and $\frac{L^2 \log L}{n} = o(\log n)$ we get

$$\lim_{n\to\infty} \mathsf{P}\left(a(\log n)\widetilde{T}_n \leq \hat{\sigma}_n(L)(y + b_1(\log n))\right) = \mathrm{e}^{-2\,\mathrm{e}^{-y}}.$$

From there

$$\lim_{n\to\infty} \mathsf{P}\left(a(\log n)\frac{\widetilde{T}_n}{\hat{\sigma}_n(L)} \leq y + b_1(\log n)\right) = \mathrm{e}^{-2\,\mathrm{e}^{-y}}.$$

Now let use take

$$\tilde{T}_n^*(L) = \frac{\tilde{T}_n}{\hat{\sigma}_n(L)} = \max_{1\leq k<n}\left\{\sqrt{\frac{n}{k(n-k)}}\cdot\frac{|S_{kn}|}{\hat{\sigma}_n(L)}\right\}. \tag{4.7}$$

Then we arrive at

$$\lim_{n\to\infty} \mathsf{P}\left(a(\log n)\tilde{T}_n^*(L) \leq y + b_1(\log n)\right) = \mathrm{e}^{-2\,\mathrm{e}^{-y}}. \tag{4.8}$$

Let us now recall the statistic T_n^* defined in (1.3), where the i.i.d. case of off-line change point detection problem was discussed and compare it to $\tilde{T}_n^*(L)$.

$$T_n^* = \max_{1\leq k<n}\left\{\sqrt{\frac{n}{k(n-k)}}\cdot\frac{|S_{kn}^*|}{\hat{\sigma}_n}\right\},$$

where S_{kn}^* stands for k-th partial sum of least squares residuals from model (1.1) with $q_n = n$. From the asymptotic results for T_n^* shown in (1.4) and for $\tilde{T}_n^*(L)$ shown in (4.8) we can observe that if we standardize these statistics we get

$$\begin{aligned} a(\log n)\tilde{T}_n^*(L) - b(\log n) &\xrightarrow{\mathcal{D}} T, \\ a(\log n)T_n^* \qquad - b(\log n) &\xrightarrow{\mathcal{D}} T, \end{aligned}$$

where T stands for a random variable with cumulative distribution function

$$F_T(y) = \mathsf{P}(T \leq y) = \mathrm{e}^{-2\,\mathrm{e}^{-y}}. \tag{4.9}$$

We can see that after we relax the assumption of i.i.d. error terms in a location model and allow them to behave as a linear process, we can still use the same statistic for testing a change in the mean and the same distribution for asymptotic approximation. The only adjustment we have to make is not to use the estimator of variance $\hat{\sigma}_n^2$, but the estimator of long-run variance $\hat{\sigma}_n^2(L)$ instead.

4.3 Chosen long-run variance estimator

The estimator $\hat{\sigma}_n^2(L)$ defined in (4.5) belongs to the class of kernel spectral density estimators which have the following general form

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \mathrm{e}^{i\omega k} \, \lambda_M(k)\hat{R}(k),$$

where $\hat{R}(\cdot)$ denotes the estimated autocovariance function, $\lambda_M : \mathbb{R} \to \mathbb{R}$ denotes the kernel function with bandwidth parameter M. $\lambda_M(t)$ can be typically written as $\lambda_M(t) = \lambda(\frac{t}{M})$, where $\lambda : \mathbb{R} \to \mathbb{R}$ satisfies $\int\limits_{-\infty}^{\infty} \lambda(s)\,\mathrm{d}s = 1$. For example the estimator $\hat{\sigma}_n^2(L)$ is a member of this class with so-called Bartlett kernel of the form

$$\lambda_B(s) = \begin{cases} 1 - |s|\,, & |s| < 1 \\ 0, & \text{otherwise,} \end{cases} \tag{4.10}$$

and the bandwidth parameter $L > 0$, as can be seen after a quick calculation and using $\hat{R}(k) = \hat{R}(-k), k = -1, \ldots, -n+1$

$$\begin{aligned} 2\pi \cdot \hat{f}_B(0) &= \sum_{k=-\infty}^{\infty} \lambda_B\left(\frac{k}{L}\right)\hat{R}(k) = \sum_{k=-\lfloor L \rfloor}^{\lfloor L \rfloor} \left(1 - \frac{|k|}{L}\right)\hat{R}(k) = \\ &= \hat{R}(0) + 2\sum_{k=1}^{\lfloor L \rfloor} \left(1 - \frac{k}{L}\right)\hat{R}(k) = \hat{\sigma}_n^2(L), \end{aligned}$$

where $\lfloor x \rfloor$ denotes the largest integer that is lower than x. There is a special kind of kernels discussed in Politis (2001) - the so-called flat-top kernels. These kernels have favorable properties in terms of the rate of convergence as described in the aforementioned article. However, these advantages come with a price. Long-run variance estimators based on flat-top kernels are not almost surely positive , but the fast rate of convergence indicates that the problematic cases may be rare. One of the simplest and most intuitive members of the flat-top kernel family is the trapezoidal kernel which is of the form

$$\lambda_T(s) = \begin{cases} 1, & \text{if } |s| \in [0, c], \\ \frac{1}{1-c}(1 - |s|), & \text{if } |s| \in [c, 1], \\ 0, & \text{otherwise,} \end{cases} \tag{4.11}$$

where $c \in (0, 1)$ is a parameter. Next picture provides a comparison betweeen trapezoidal and Bartlett kernels and explains where did the name “trapezoidal” come from.

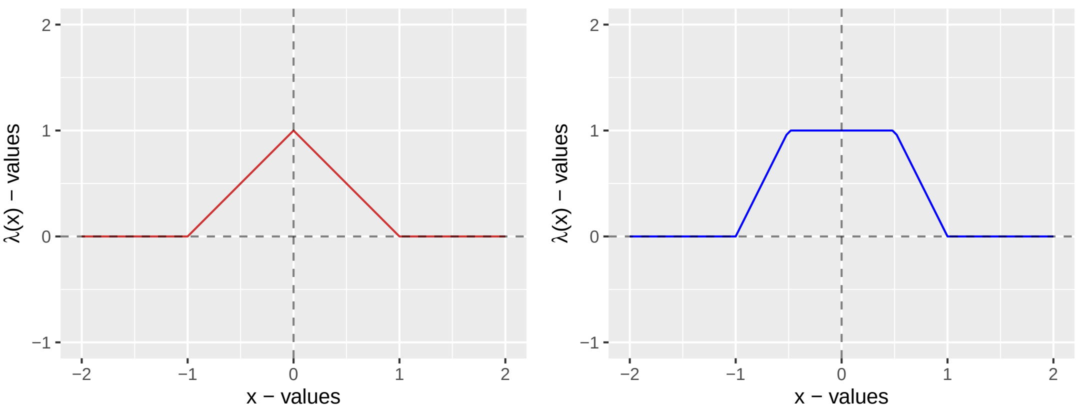


Figure 4.1: Bartlett and trapezoidal kernel ($c = \frac{1}{2}$) functions

We can see the difference in the weights the kernels give to estimates $\hat{R}(k)$ with increasing k. Estimator that uses trapezoidal kernel gives the same weight to all $\hat{R}(k)$ up to $M \cdot c$, then the weights start to decay linearly and reach zero at point M. On the other hand, estimator using Bartlett kernel gives the highest weight to $\hat{R}(0)$ and the weights of autocovariance function estimates decrease linearly with lag until they reach zero at point M. The idea behind the term "bandwidth" for M is apparent. Due to $\lambda(x) = 0, x \geq M$ for both kernels, we have that only $\hat{R}(k), k < M$ affect the resulting estimate. We can alter the behavior of the trapezoidal kernel estimator by altering the constant c. However, the most commonly used is $c = \frac{1}{2}$.

The weakness of trapezoidal kernel when compared to other members of the flat-top family lies in the fact that it is not differentiable, on the other hand it is simple and intuitive. We chose to estimate σ^{*2} using the trapezoidal kernel spectral density estimator (with constant $c = \frac{1}{2}$)

$$\hat{\sigma}_n^2(M) = \sum_{k=-n+1}^{n-1} \lambda_T \left(\frac{k}{M} \right) \hat{R}(k), \tag{4.12}$$

where $\hat{R}(k)$ is defined in (4.6).

The problem that flat-top kernel estimators in general suffer is that they sometimes provide negative variance estimates, i.e. the estimator $\hat{\sigma}_n^2(M)$ is not almost surely non-negative as opposed to for example Bartlett kernel estimator. Nevertheless, the fast rate of convergence to a positive value indicates that these cases may be rare. However, if we want to use this estimator in finite-sample problems we need to modify it. This problem was analyzed in Politis (2011), where it was suggested to modify the estimator as follows

$$\hat{\sigma}_n^{+2}(M) = \max \left\{ \xi_n, \hat{\sigma}_n^2(M) \right\}, \tag{4.13}$$

where $\xi_n > 0$ is a chosen constant which can depend on n. In the aforementioned article it was proven that if ξ_n is chosen wisely then the estimator maintains its fast convergence while solving the problem of non-positiveness. The constant ξ_n should depend on the sample size n and the suggested choice in Politis (2011) was to take

$$\xi_n = \frac{2\pi}{n^a}, \quad 1 \leq a \leq 2.$$

4.3.1 Automatic bandwidth selection

We already decided on the chosen kernel of our spectral density estimator. What remains is to choose a proper bandwidth M. We will use an automatic bandwidth selection procedure described in Politis (2003), which is designed to find bandwidth parameter specifically for flat-top kernels. It is able to estimate the optimal value for parameter M in the most typical cases of dependency structure of the underlying process - polynomial and exponential decay of the autocovariance function and the case when there is finite q such that $R(k)$ is equal to zero for lags higher than q. The following theorem proven in Politis (2003) gives the optimal (with respect to minimization of the large-sample MSE of $\tilde{f}_l(\omega)$) behavior of M when estimating infinite sums of the type

$$f_l(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} |k|^l \, \mathrm{e}^{ik\omega} \, R(k)$$

using flat-top kernel estimates

$$\tilde{f}_l(\omega) = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} |k|^l \, \lambda\left(\frac{k}{M}\right) \mathrm{e}^{ik\omega} \, \hat{R}(k).$$

Theorem 2. *Assume conditions strong enough to ensure that*

$$var\left(\tilde{f}_l(\omega)\right) = O\left(\frac{M}{n}\right) \tag{4.14}$$

1. *Assume that* $\sum\limits_{t=-\infty}^{\infty} |t|^{(l+r)} \, |R(t)| < \infty$ *for some positive integer* r*; then letting* M *proportional to* $n^{\frac{1}{(2r+1)}}$ *yields*

$$\tilde{f}_l(\omega) = f_l(\omega) + O_P\left(n^{\frac{-r}{(2r+1)}}\right).$$

2. *If* $|R(t)| \leq C\,\mathrm{e}^{-at}$ *for some constants* $a, C > 0$*, then letting* $M \sim A \log n$*, for some appropriate constant* A*, yields*

$$\tilde{f}_l(\omega) = f_l(\omega) + O_P\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right),$$

where $A \sim B$ *stands for* $\frac{A}{B} \to 1$.

3. *If* $R(t) = 0$ *for* $|t| >$ *some* q*, then letting* $M = 2q$ *yields*

$$\tilde{f}_l(\omega) = f_l(\omega) + O_P\left(\frac{1}{\sqrt{n}}\right).$$

Note. There exist different sets of conditions for (4.14); first, let us define the α-mixing coefficients $\alpha_X(k)$ as follows: let $\mathcal{F}_k^l$ be the σ-algebra generated by $\{X_t, k \leq t \leq l\}$ and define $\alpha_X(k) = \sup_n \sup_{A,B} |\mathsf{P}(A \cap B) - \mathsf{P}(A)\,\mathsf{P}(B)|$, where A and B vary over the σ-fields $\mathcal{F}_{-\infty}^n$ and $\mathcal{F}_{n+k}^\infty$, respectively. If we assume $\mathsf{E}\,|X_t|^{6+\delta} < \infty$ and $\sum_{k=1}^{\infty} k^2(\alpha_X(k))^{\frac{\delta}{6+\delta}} < \infty$ for some $\delta > 0$, then (4.14) is fulfilled, for further information see Rosenblatt (1956).

Besides the favorable asymptotic properties and speed of convergence another reason for using the flat-top kernels is that choosing their bandwidth in practice is intuitive and doable by a simple inspection of the correlogram, i.e. the plot of $\hat{R}(k)$ against k. In Politis (2003) it was suggested to look for a point $\hat{m}$ after which the correlogram "seems negligible" and then to take $M = 2\hat{m}$. Correct description of the procedure will be described in the next paragraph. In Politis (2003) it was proven that this simple approach automatically captures the correct magnitude for M, thus adapting to the three cases described in Theorem 2.

Empirical rule of picking M: Let $\rho(k) = \frac{R(k)}{R(0)}$ and $\hat{\rho}(k) = \frac{\hat{R}(k)}{\hat{R}(0)}$, where $R(\cdot)$ denotes autocovariance function and $\hat{R}(k)$ its estimator as defined in (4.6). Let $\hat{m}$ be the smallest positive integer such that $|\hat{\rho}(\hat{m}+k)| < c\sqrt{\frac{\log_{10} n}{n}}$, for $k = 1, \ldots, K_n$, where $c > 0$ is a fixed constant, and K_n is positive, non-decreasing integer-valued function of n such that $K_n = o(\log n)$ and $\log_{10}$ denotes logarithm with base 10. Then, let $\hat{M} = 2\hat{m}$.

Because $\hat{\rho}(k) = 0, |k| \geq n$, the above minimization problem is always well-defined, although a case where $\hat{m}$ and $\hat{M}$ turn out comparable to n deserves further scrutiny, as is well-known we need $\hat{m}$ and $\hat{M}$ to be of smaller order than n in order to have estimators with small variance.

The constants c and K_n are the practitioner's choice, indeed any values for $c > 0$ and $1 \leq K_n \leq n$ would work for the following asymptotic results albeit leading to very different finite-sample performances. The guidance on practically useful choices for c and K_n in Politis (2003) comes from the comparison with the construction of confidence intervals/hypothesis tests for the autocorrelations. They recommend on taking K_n to be about 5 and c to a value around 2 so that our empirical rule would roughly correspond to 95% simultaneous intervals for $\rho(\hat{m}+k)$ with $k = 1, \ldots, K_n$ by Bonferroni's inequality. It is important to point out that this rule-of-thumb is *only applicable with the flat-top lag windows.* It can not be applied to traditional lag windows. The performance of the empirical rule for picking M is quantified in the following theorem proven in Politis (2003), where also the sufficient conditions described in the following note were mentioned.

Theorem 3. *Assume conditions strong enough to ensure that for all finite m,*

$$\max_{k=1,\ldots,m} |\hat{\rho}(s+k) - \rho(s+k)| = O_P\left(\frac{1}{\sqrt{n}}\right) \tag{4.15}$$

uniformly in s, and

$$\max_{k=1,\ldots,n-1} |\hat{\rho}(k) - \rho(k)| = O_P\left(\frac{\log n}{\sqrt{n}}\right) \tag{4.16}$$

Also assume $|R(k)| > 0$ for all $k \leq$ some k_0.

1. *Assume that $R(k) = Ck^{-d}$ for $k > k_0$, and for some $C > 0$, and a positive integer d. Then,*

$$\hat{M} \overset{P}{\sim} \frac{A_1 n^{1/2d}}{(\log n)^{1/2d}}$$

for some positive constant A_1, where $A \overset{P}{\sim} B$ means $\frac{A}{B} \to 1$ in probability.

2. *Assume that $R(k) = C\xi^k$ for $k > k_0$, where $C > 0$, and $|\xi| < 1$ are some constants. Then,*

$$\hat{M} \overset{P}{\sim} A_2 \cdot \log n$$

where $A_2 = -1/\log|\xi|$.

3. *If $R(k) = 0$ for all $k > q \equiv k_0$, but $R(q) \neq 0$, then*

$$\hat{M} = 2q + o_P(1)$$

Note. There exist different sets of conditions for (4.15), see Brockwell and Davis (2013) or Romano and Thombs (1996). As a matter of fact, under further regularity conditions, the process $\sqrt{n}(\hat{\rho}(\cdot) - \rho(\cdot))$ is asymptotically Gaussian with autocovariance tending to zero, consequently (4.16) would follow from theory of extremes of dependent sequences.

Comparing the empirical rule $\hat{M}$ to the theoretically optimal M given in Theorem 2 we see that $\hat{M}$ manages to capture exactly the theoretically optimal rate in cases (ii) and (iii) of Theorem 3. In case (i) $\hat{M}$ increases essentially as a power of n since the $2d$-th root of the logarithm changes in a very slow rate with n. The empirical exponent $1/2d$ is slightly smaller than the theoretically optimal, but the difference is small and becomes even smaller for large d.

Thus, $\hat{M}$ automatically adapts to the underlying rate of decay of the autocorrelation function, switching between the polynomial, logarithmic and constant rates that are optimal respectively in the three cases of Theorem 2.

Now we have an intuitive estimate $\hat{M}_A$, based on the introduced procedure, which we can plug into our flat-top kernel estimator to get $\sigma_n^2(\hat{M}_A)$. If we put this together with (4.13) and use $a = 1$ we can take

$$\hat{\sigma}_{A,n}^{+2} = \max\left\{\frac{2\pi}{n}, \hat{\sigma}_n^2(\hat{M}_A)\right\} \tag{4.17}$$

and we get the test statistic of the form

$$T_n^A = \max_{1 \leq k < n}\left\{\sqrt{\frac{n}{k(n-k)}} \cdot \frac{|S_{kn}|}{\hat{\sigma}_{A,n}^{+}}\right\}. \tag{4.18}$$

It should be mentioned that similar automatic choices should be approached with caution and should be validated by the practitioner. In situations where it is not possible, like for example in bootstrap procedures, artificial bounds are used to control at least the most extreme cases.

4.4 Applied block bootstrap procedures

We decided to use block bootstrap and dependent wild bootstrap procedures, which were introduced in the third chapter, for the simulations performed on the upcoming pages. We will make use of a modification of block bootstrap procedures introduced in Politis and Romano (1992) called circular bootstrap.

It consists of "wrapping" the data around in a circle before blocking them. The advantage lies in the fact that resulting bootstrap series is automatically centered around the sample mean. Another reason to use the circular bootstrap is that Politis and White (2004) introduced an automatic procedure for estimating the optimal block length for circular block bootstrap methods. Nevertheless, circular and non-circular block bootstrap procedures are asymptotically equivalent.

To achieve the "wrapping" we first define

$$
\begin{aligned}
X_i &= X_{(i \bmod n)}, \quad i > n, \\
X_0 &= X_n,
\end{aligned}
$$

where $i \bmod n$ denotes "modulo n". Now we can define block $(X_i, X_{i+1}, \ldots, X_{i+l-1})$ for any $i = 1, \ldots, n$ and any block length $l > 0$. We can use these blocks in our algorithms. Before we discuss block length selection let us go over a general circular block bootstrap method:

1. We start by "wrapping" the data
$$
\begin{aligned}
X_i &= X_{(i \bmod n)}, \quad i > n, \\
X_0 &= X_n.
\end{aligned}
$$

2. We draw $i_0, i_1, \ldots$ i.i.d. from uniform distribution on the set $\{1, \ldots, n\}$ - the starting points of our blocks.

3. We draw $b_0, b_1, \ldots$ i.i.d. from some distribution $F_b(\cdot)$ that depends on parameter b (which may depend on n) - the block sizes.

4. We construct a bootstrap pseudo-series $X_1^*, X_2^*, \ldots, X_n^*$ as follows. For $m = 0, 1, \ldots,$ let
$$\mathbf{X}_m^* = (X_{i_m}, \ldots, X_{i_m+b_m-1})$$
and $(X_1^*, \ldots, X_n^*)$ corresponds to the first n elements of the compounded vector $(\mathbf{X}_0^*, \mathbf{X}_1^*, \ldots)$.

   This procedure defines a probability measure (conditional on the data X_1, $X_2, \ldots, X_n$) denoted as P^*; expectation and variance with respect to P^* are denoted as E^* and var* respectively.

5. We can then construct the bootstrap sample mean $\bar{X}_n^* = \frac{1}{n}\sum_{i=1}^{n} X_i^*$. The corresponding estimate of the asymptotic variance of the sample mean is then given by $\mathrm{var}^*(\sqrt{n}\bar{X}_n^*)$.

The cases of fixed and random block length circular bootstrap correspond to situations:

(A) F_b is a unit mass on the positive integer b; this corresponds to circular block bootstrap with fixed block length b which will be denoted as *circular bootstrap* (CB) further on and the corresponding estimate of $\sigma^{*2} = \sum_{k=-\infty}^{\infty} R(k)$ will be denoted as $\hat{\sigma}^2_{b,CB}$.

(B) F_b is a Geometric distribution with mean equal to the real number b; this corresponds to circular block bootstrap with random block length which will be denoted as *stationary bootstrap* (SB) further on and the corresponding estimate of σ^{*2} will be denoted as $\hat{\sigma}^2_{b,SB}$.

The next theorem states the basis for the optimization of block lengths in Politis and White (2004) and was proven in Lahiri (1999).

Theorem 4. *Assume a stationary process $\{X_t\}$ with autocovariance function $R(\cdot)$. Let $\{X_t\}$ be strong mixing, i.e. $\alpha_X(k) \to 0, k \to \infty$, where $\alpha_X(k)$ are defined as in Note 4.3.1. Furthermore let $\mathsf{E}\,|X_t|^{6+\delta} < \infty$, and $\sum_{k=1}^{\infty} k^2(\alpha_X(k))^{\frac{\delta}{6+\delta}} < \infty$ for some $\delta < 0$ and let us denote*

$$\sigma^{*2} = \sum_{k=-\infty}^{\infty} R(k)$$

and

$$g(\omega) = \sum_{s=-\infty}^{\infty} R(s)\cos(\omega s).$$

If $b \to \infty$ as $n \to \infty$, but with $b = o(\sqrt{n})$, then we have:

$$\begin{aligned}
bias(\hat{\sigma}^2_{b,CB}) &= \mathsf{E}\,\hat{\sigma}^2_{b,CB} - \sigma^{*2} = -\frac{1}{b}G + o(b^{-1}),\\
var(\hat{\sigma}^2_{b,CB}) &= \frac{b}{n}D_{CB} + o\left(\frac{b}{n}\right);\\
bias(\hat{\sigma}^2_{b,SB}) &= \mathsf{E}\,\hat{\sigma}^2_{b,SB} - \sigma^{*2} = -\frac{1}{b}G + o(b^{-1}),\\
var(\hat{\sigma}^2_{b,SB}) &= \frac{b}{n}D_{SB} + o\left(\frac{b}{n}\right);
\end{aligned}$$

in the above, $D_{CB} = \frac{4}{3}g^2(0)$, $D_{SB} = (4g^2(0) + \frac{2}{\pi}\int_{-\infty}^{\infty}(1+\cos\omega)g^2(\omega)\,\mathrm{d}\omega)$, $G = \sum_{k=-\infty}^{\infty}|k|\,R(k)$.

From the above theorem it is apparent that SB is less accurate than CB for estimating σ^{*2}. Although the two methods have similar bias, the SB has higher variance due to the additional randomization involved in the random block sizes.

4.4.1 Block length selection for the stationary bootstrap

From Theorem 4 it follows that for the *stationary bootstrap* we have:

$$\mathrm{MSE}(\hat{\sigma}^2_{b,SB}) = \frac{G^2}{b^2} + D_{SB}\frac{b}{n} + o(b^{-2}) + o\left(\frac{b}{n}\right)$$

It now follows that the large-sample $\mathrm{MSE}(\hat{\sigma}^2_{b,SB})$ is minimized if we choose

$$b_{opt,SB} = \left(\frac{2G^2}{D_{SB}}\right)^{\frac{1}{3}} \cdot n^{\frac{1}{3}}.$$

Using the optimal block size $b_{opt,SB}$ we achieve the optimal MSE, which is given by

$$\mathrm{MSE}_{opt,SB} \sim \frac{3}{2^{\frac{2}{3}}} \cdot \frac{G^{\frac{2}{3}} D_{SB}^{\frac{2}{3}}}{n^{\frac{2}{3}}}.$$

The quantities G, D_{SB} and D_{CB} involve unknown parameters $\sum\limits_{k=-\infty}^{\infty} |k|\, R(k)$, $\sigma^{*2} = \sum\limits_{k=-\infty}^{\infty} R(k) = g(0)$ and $\frac{1}{\pi} \int\limits_{-\pi}^{\pi} (1 + \cos\omega) g^2(\omega)$. These must be estimated in order to obtain any practical use for the procedure.

In the aforementioned article Politis and White (2004) flat-top kernel estimation is recommended. Thus, we estimate $\sum\limits_{k=-\infty}^{\infty} |k|\, R(k)$ by $\sum\limits_{k=-\infty}^{\infty} \lambda_T(\frac{k}{M})\, |k|\, \hat{R}(k)$, where $\lambda_T(\cdot)$ denotes trapezoidal flat-top kernel as in (4.11).

Similarly, we estimate $g(\omega) = \sum\limits_{k=-\infty}^{\infty} R(k) \cos(\omega k)$ by $\hat{g}(\omega) = \sum\limits_{k=-n+1}^{n-1} \lambda_T(\frac{k}{M}) \cdot$ $\cdot \hat{R}(k) \cos(\omega k)$. Plugging the two estimators in the expression for G and D_{SB} we arrive at the estimators

$$\hat{G} = \sum_{k=-n+1}^{n-1} \lambda_T\left(\frac{k}{M}\right) |k|\, \hat{R}(k),$$

$$\hat{D}_{SB} = \left(4\hat{g}^2(0) + \frac{2}{\pi}\int\limits_{-\pi}^{\pi} (1 + \cos\omega)\hat{g}^2(\omega)\, \mathrm{d}\omega\right).$$

Thus, the estimator introduced in Politis and White (2004) is then of the form

$$\hat{b}_{opt,SB} = \left(\frac{2\hat{G}^2}{\hat{D}_{SB}}\right)^{\frac{1}{3}} \cdot n^{\frac{1}{3}}. \tag{4.19}$$

Asymptotic performance of the suggested $\hat{b}_{opt,SB}$ is given by the following theorem proven in Politis and White (2004).

Theorem 5. *Assume the conditions of Theorem 4 hold.*

1. *Assume that* $\sum_{k=-\infty}^{\infty} |k|^{r+1} |R(k)| < \infty$ *for some positive integer* r*; then taking* M *proportional to* $n^{1/(2r+1)}$ *yields*

$$\hat{b}_{opt,SB} = b_{opt,SB}\left(1 + O_P(n^{\frac{-r}{2r+1}})\right).$$

2. *If* $R(k)$ *has an exponential decay, then taking* $M \sim A \log n$*, for some given non-negative* A*, yields*

$$\hat{b}_{opt,SB} = b_{opt,SB}\left(1 + O_P\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right)\right).$$

3. *If* $R(k) = 0$ *for* $|k|$ *greater than some integer* q*, then taking* $M = 2q$ *yields*

$$\hat{b}_{opt,SB} = b_{opt,SB}\left(1 + O_P\left(\frac{1}{\sqrt{n}}\right)\right).$$

Using the flat-top estimator has an advantage in the fast convergence and also in a simple way of choosing M which is described sooner in this section 4.3.1 and involves identifying a point $\hat{m}$ after which the estimated correlogram $\hat{R}(\cdot)$ appears negligible, i.e. $\hat{R}(k) \simeq 0, k > \hat{m}$. Then we take $\hat{M} = 2\hat{m}$. Perhaps the most interesting feature about this rule is its adaptivity to different correlation structures. It is an *omnibus* rule-of-thumb that automatically gives good bandwidth choices for the flat-top kernel without having to prespecify the correlation structure. Again it is important to point out that it is *only applicable with the flat-top lag windows* and can not be applied to the more classical lag windows.

4.4.2 Block length selection for the fixed block length bootstrap

Similarly to the stationary case, for the *circular bootstrap* we have:

$$\text{MSE}(\hat{\sigma}^2_{b,CB}) = \frac{G^2}{b^2} + D_{CB}\frac{b}{n} + o(b^{-2}) + o\left(\frac{b}{n}\right)$$

It now follows that the large-sample $\text{MSE}(\hat{\sigma}^2_{b,SB})$ is minimized if we choose

$$b_{opt,SB} = \left[\left(\frac{2G^2}{D_{CB}}\right)^{\frac{1}{3}} \cdot n^{\frac{1}{3}}\right],$$

where $[x]$ denotes the closest integer to the real number x. Using the optimal block size $b_{opt,CB}$ we achieve the optimal MSE, which is given by

$$\text{MSE}_{opt,CB} \sim \frac{3}{2^{\frac{2}{3}}} \cdot \frac{G^{\frac{2}{3}} D_{CB}^{\frac{2}{3}}}{n^{\frac{2}{3}}}.$$

Plugging in our estimator $\hat{g}$ for g in the expression for D_{CB} we get

$$\hat{D}_{CB} = \frac{4}{3} \cdot \hat{g}^2(0).$$

Estimating G by $\hat{G}$, we are led to the following optimal block size estimator:

$$\hat{b}_{opt,CB} = \left[\left(\frac{2\hat{G}^2}{\hat{D}_{CB}} \right)^{\frac{1}{3}} \cdot n^{\frac{1}{3}} \right]. \tag{4.20}$$

The behavior of $\hat{b}_{opt,CB}$ is similar to $\hat{b}_{opt,SB}$ as the following theorem states.

Theorem 6. *Assume the conditions of Theorem 4 hold.*

1. *Assume that* $\sum_{k=-\infty}^{\infty} |k|^{r+1} |R(k)| < \infty$ *for some positive integer* r*; then taking* M *proportional to* $n^{1/(2r+1)}$ *yields*

$$\hat{b}_{opt,CB} = b_{opt,CB}(1 + O_P(n^{\frac{-r}{2r+1}})).$$

2. *If* $R(k)$ *has an exponential decay, then taking* $M \sim A \log n$*, for some given non-negative* A*, yields*

$$\hat{b}_{opt,CB} = b_{opt,CB} \left(1 + O_P \left(\frac{\sqrt{\log n}}{\sqrt{n}} \right) \right).$$

3. *If* $R(k) = 0$ *for* $|k|$ *greater than some integer* q*, then taking* $M = 2q$ *yields*

$$\hat{b}_{opt,CB} = b_{opt,CB} \left(1 + O_P \left(\frac{1}{\sqrt{n}} \right) \right).$$

Again we will use the $\hat{M} = 2\hat{m}$ rule-of-thumb mentioned earlier in section 4.3.1 in the estimates $\hat{G}, \hat{g}(0)$.

Chapter 5

Asymptotic results

We assume the location model described by (4.1) and we are still interested in testing the null hypothesis

$$H_0 : q_n = n \text{ against } H_1 : q_n < n, \tag{5.1}$$

i.e. whether there occurred a change in the mean of the series or not. Denote

$$\tilde{T}_n = \max_{1 \le k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \cdot |S_{kn}| \right\}.$$

5.1 Asymptotic distribution

Based on (4.8) under the null hypothesis, assumptions (4.1), (4.2) and (4.3) and assuming $\frac{L^2}{\log n} \to \infty$ and $\frac{L^2 \log L}{n} = o(\log n)$ we have for any $y \in \mathbb{R}$

$$\lim_{n\to\infty} \mathsf{P}\left(a(\log n) \cdot \frac{\tilde{T}_n}{\hat{\sigma}_n(L)} \le b_1(\log n) + y \right) = \mathrm{e}^{-2\,\mathrm{e}^{-y}},$$

where $a(x) = \sqrt{2\log x}$ and $b_p(x) = 2\log x + \frac{p}{2}\log\log x - \log\left(\Gamma\left(\frac{p}{2}\right)\right), x > 1$. According to Antoch et al. (1997) we were able to replace σ^* in Theorem 1 with $\hat{\sigma}_n(L)$. We will now show that it remains valid even when σ^* is replaced by the estimate based on the trapezoidal kernel $\hat{\sigma}_n(M)$, if M is chosen correctly.

To show this assertion we will use Theorem 1.1. from Kirch (2007). It in fact shows, that under the assumptions from [b.1.] and H_0

$$\lim_{n\to\infty} \mathsf{P}\left(a(\log n) \cdot \frac{\tilde{T}_n}{\hat{s}_n} - b_1(\log n) \le y \right) = \mathrm{e}^{-2\,\mathrm{e}^{-y}}$$

holds for any estimator $\hat{s}_n^2$ of long-run variance satisfying

$$\hat{s}_n - \sigma^* = o_{\mathsf{P}}((\log\log n)^{-1}). \tag{5.2}$$

Let us now recall Theorem 2. We have

$$\begin{aligned} l &= 0, \\ \hat{\sigma}_n^2(M) &= 2\pi \cdot \tilde{f}_0(0), \\ \sigma^{*2} &= 2\pi \cdot f_0(0). \end{aligned}$$

In our location model (4.8) let us assume (4.1), (4.2) and (4.3) and let the assumptions of Theorem 2 be satisfied. Furthermore let the null hypothesis hold, the autocovariance function $R(\cdot)$ of $\{y(t)\}$ be as in (i), (ii) or (iii) of Theorem 2 and M be of the corresponding magnitude described in the aforementioned theorem. Then there exists $\gamma > 0$ such that

$$\hat{\sigma}_n^2(M) - \sigma^{*2} = o_{\mathsf{P}}(n^{-\gamma}), \quad n \to \infty.$$

We will make use of the operations with o_{P} and the Taylor expansion of function $c(x) = \sqrt{1+x}$ around 0 which shows

$$c(x) = \sqrt{1+x} = 1 + \frac{1}{2} \cdot x + o(x), x \to 0.$$

Using these results we obtain

$$\begin{aligned} \hat{\sigma}_n(M) - \sigma^* = \sigma^* \left(\frac{\hat{\sigma}_n(M)}{\sigma^*} - 1 \right) &= \sigma^* \left(\sqrt{1 + \frac{\hat{\sigma}_n^2(M) - \sigma^{*2}}{\sigma^{*2}}} - 1 \right) = \\ &= \sigma^* \left(1 + \frac{1}{2} \cdot \frac{\hat{\sigma}_n^2(M) - \sigma^{*2}}{\sigma^{*2}} + o_{\mathsf{P}}(n^{-\gamma}) - 1 \right) = \\ &= \frac{1}{2\sigma^*} \left(\hat{\sigma}_n^2(M) - \sigma^{*2} \right) + o_{\mathsf{P}}(n^{-\gamma}) = o_{\mathsf{P}}(n^{-\gamma}), n \to \infty. \end{aligned}$$

Plugging this result back to Theorem 1.1. from Kirch (2007) we get the assertion

$$\lim_{n\to\infty} \mathsf{P} \left(a(\log n) \cdot \frac{\tilde{T}_n}{\hat{\sigma}_n(M)} - b_1(\log n) \le y \right) = \mathrm{e}^{-2\,\mathrm{e}^{-y}}.$$

Note. We assumed that our bandwidth M was of the theoretically optimal magnitude described in Theorem 2. However, according to Theorem 3 the automatic bandwidth choice $\hat{M}_A$ discussed in 4.3.1 automatically adapts to the underlying rate of decay and manages to capture the desired magnitude very well.

We have

$$\lim_{n\to\infty} \mathsf{P} \left(a(\log n) \cdot \frac{\tilde{T}_n}{\hat{\sigma}_n(M)} - b_1(\log n) \le -\log\left(-\frac{1}{2} \log t \right) \right) = t.$$

The previous result suggests a way of standardizing our test statistic as

$$T_n^S = a(\log n) \cdot T_n^A - b_1(\log n), \tag{5.3}$$

Let us now take a look on random variable T with limiting distribution having $F_T = \mathrm{e}^{-2\,\mathrm{e}^{-y}}$. After we change the notation we can derive

$$
\begin{aligned}
t &= \mathrm{e}^{-2\,\mathrm{e}^{-y}}, \\
-\frac{1}{2}\log t &= \mathrm{e}^{-y}, \\
-\log\left(-\frac{1}{2}\log t\right) &= y.
\end{aligned}
$$

then the quantiles of this distribution, which will be used as asymptotic critical values in our simulations in the following sections, are of the form

$$C_{as}(t) = -\log\left(-\frac{1}{2}\log t\right).$$

These values are summarized in the following table.

t	0.9	0.95	0.975	0.99	0.999
$C_{as}(t)$	2.94	3.66	4.37	5.29	7.60

Table 5.1: Quantiles of limiting distribution with CDF $F_T(\cdot)$ as in (4.9)

Let us now define a test procedure of our null hypothesis based on asymptotic critical values. Assume we have observations $y(1), \ldots, y(n)$ from model (4.1) and we are interested in testing (5.3) on significance level $\alpha \in (0, 1)$. We calculate T_n^S based on our data and reject H_0 if and only if

$$T_n^S > C_{as}(1-\alpha). \tag{5.4}$$

The following chapter aims on presenting the results of the performed simulation study on the location model and the reader's introduction into the bootstrap procedures implementation.

Chapter 6

Simulation study in location model

In our simulation study we will focus on the model described in (4.1), where the error terms $\{e(i)\}_{i=-\infty}^{\infty}$ will behave as a typical example of a linear process - autoregressive AR(1) process. They satisfy

$$e(t) = \rho \cdot e(t-1) + \epsilon_t, \quad t \in \mathbb{Z}.$$

When simulating samples from AR(1) process it is impossible to allow for the infinite history of dependence. Therefore, the simulations of length n from this process were obtained as follows.

$$\tilde{e}(0) = 0,$$
$$\tilde{e}(t) = \rho \cdot \tilde{e}(t-1) + \epsilon_t, \quad t = 1, 2, \dots,$$

where $\rho \in (-1, 1)$ is a parameter and $\{\epsilon_s\}_1^{\infty}$ are normally distributed i.i.d. random variables having zero mean and unit variance. Our sample is then taken as

$$e(t) = \tilde{e}(50 + t), t = 1 \dots, n$$

Under the null hypothesis our model (4.1) depends on parameters $\rho \in (-1, 1)$, $n \in \mathbb{N}$ and $\mu \in \mathbb{R}$. In the following simulation μ was set to be equal to zero, ρ ranges in set $\{-0.5, -0.3, 0, 0.3, 0.5, 0.7\}$ and n ranges in set $\{100, 200, 500\}$. All of the simulations were performed using the **R** software and the source code can be found as an attachment to this thesis.

For each combination of ρ and n the simulation algorithm was set as follows:

1. A sample $y(1), \dots, y(n)$ of length n was taken from model (4.1) under the null hypothesis, i.e. from AR(1) model.
2. The estimated optimal bandwidth $\hat{M}_A$ was calculated from our sample according to the empirical rule-of-thumb described in section 4.3.1.
3. The estimate $\hat{\sigma}_{A,n}^{+2}$ based on trapezoidal flat-top kernel with bandwidth $\hat{M}_A$ was calculated as described in (4.17).

4. The value of the standardized test statistic T_n^S was calculated as described in (5.3).

This procedure was repeated 10000 times and sample quantiles of the resulting set of observations of T_n^S are reported in the following tables.

			n = 100		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.87	2.27	2.72	4.34	22.77
0.5	2.29	2.73	3.14	3.75	13.92
0.3	2.20	2.63	3.05	3.65	10.24
0	2.29	2.72	3.13	3.65	6.19
-0.3	4.59	5.82	7.17	9.24	14.26
-0.5	10.31	12.25	14.31	17.10	23.25

			n = 200		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.83	2.21	2.55	2.96	8.13
0.5	2.24	2.67	3.05	3.57	5.24
0.3	2.34	2.73	3.11	3.61	4.63
0	2.29	2.71	3.08	3.50	4.54
-0.3	4.03	4.82	5.60	6.70	11.17
-0.5	10.00	13.66	17.01	20.85	29.90

			n = 500		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.85	2.21	2.57	3.03	3.82
0.5	2.17	2.63	3.04	3.49	4.50
0.3	2.38	2.86	3.30	3.80	4.84
0	2.33	2.78	3.23	3.77	4.72
-0.3	3.69	4.31	4.91	5.63	7.50
-0.5	3.82	4.82	5.80	7.16	11.27

Table 6.1: Tables displaying the sample quantiles of simulated T_n^S statistics

In Table 6.1 we can see the results of our simulations. First of all, the high values of the 99.9-th percentile are caused by the fact that the original trapezoidal kernel estimator is not almost surely positive and we had to modify it to get $\hat{\sigma}_{A,n}^{+2}$ which underestimates standard deviation in case we hit the low bound. We can see that as was mentioned with rising n these cases become more and more rare. Interestingly these cases occur mostly in the case when $\rho = -0.5$ which can be traced to the fact that here $\sigma^* = (1-\rho)^{-1}$ is closest to zero.

The Table 6.2 summarizes the behavior of our automatic bandwidth choice. We can observe that the automatic choice gives in general higher values for higher values of ρ, which is expected. We do need broader bandwidths to capture lingering dependency. For all sample sizes the bandwidths are smallest for $\rho = 0$, i.e. i.i.d. case.

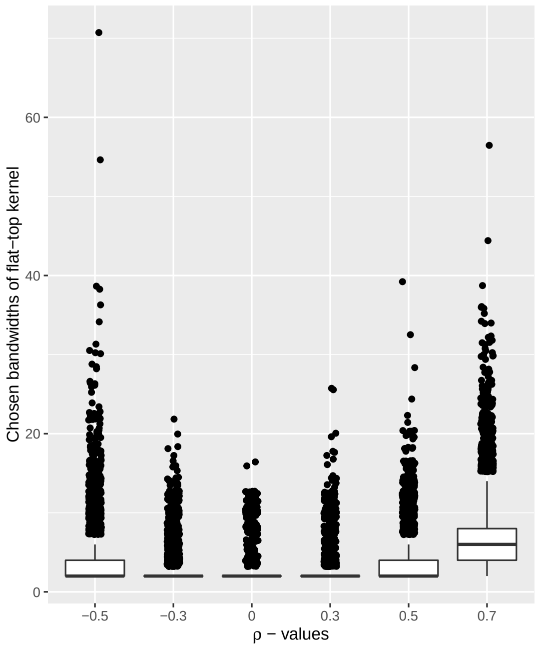
			n = 100				
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	2.00	4.00	6.00	8.00	14.00	56.00	6.25
0.5	2.00	2.00	2.00	4.00	6.00	40.00	2.99
0.3	2.00	2.00	2.00	2.00	2.00	26.00	2.21
0	2.00	2.00	2.00	2.00	2.00	16.00	2.10
-0.3	2.00	2.00	2.00	2.00	2.00	22.00	2.24
-0.5	2.00	2.00	2.00	4.00	8.00	70.00	3.31

			n = 200				
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	2.00	6.00	6.00	10.00	16.00	56.00	8.18
0.5	2.00	2.00	4.00	4.00	6.00	34.00	3.71
0.3	2.00	2.00	2.00	2.00	2.00	18.00	2.17
0	2.00	2.00	2.00	2.00	2.00	16.00	2.06
-0.3	2.00	2.00	2.00	2.00	4.00	28.00	2.21
-0.5	2.00	2.00	4.00	4.00	8.00	36.00	3.94

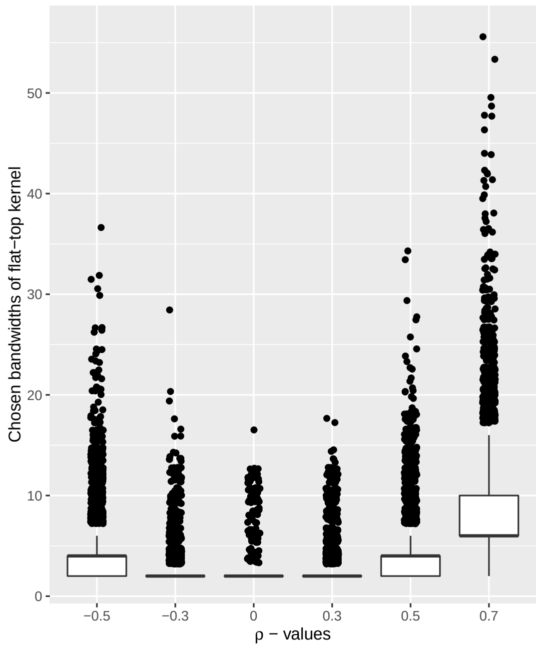
			n = 500				
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	4.00	8.00	10.00	12.00	18.00	62.00	10.39
0.5	2.00	4.00	4.00	6.00	8.00	38.00	4.84
0.3	2.00	2.00	2.00	2.00	4.00	14.00	2.26
0	2.00	2.00	2.00	2.00	2.00	12.00	2.03
-0.3	2.00	2.00	2.00	2.00	4.00	16.00	2.29
-0.5	2.00	4.00	4.00	6.00	8.00	32.00	4.98

Table 6.2: Tables displaying the behavior of the automatic bandwidth choice described in section 4.3.1

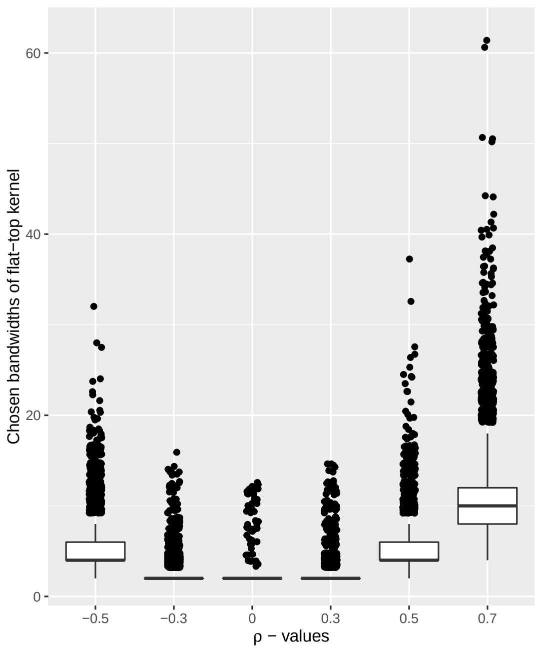
The results displayed in Table 6.2 are illustrated using boxplot Figure 6.1. The upper bound of the box is equal to the third quartile, the lower bound to the first quartile and the line in between equals to the median. The upper whisker extends from the upper bound to the highest value that is within $1.5 \cdot IQR$ of the hinge, where IQR is the inter-quartile range, or distance between the first and third quartiles. The lower whisker extends from the lower bound to the lowest value within $1.5 \cdot IQR$ of the hinge. Data beyond the end of the whiskers are outliers and are plotted as points. In most cases the empirical quartiles are too close to each other and the boxes appear as lines. However, it is visible that in general higher bandwidths appear for higher absolute values of ρ.



(a) n = 100



(b) n = 200



(c) n = 500

Figure 6.1: Boxplots illustrating the simulated automatic bandwidth choices for varying samples sizes n

Table 6.3 illustrates the behavior of our trapezoidal spectral kernel estimator of σ^{*2} compared to its actual value, which can be found in the last column. We can see that with increasing sample size, the estimates in the mean get closer to the actual value.

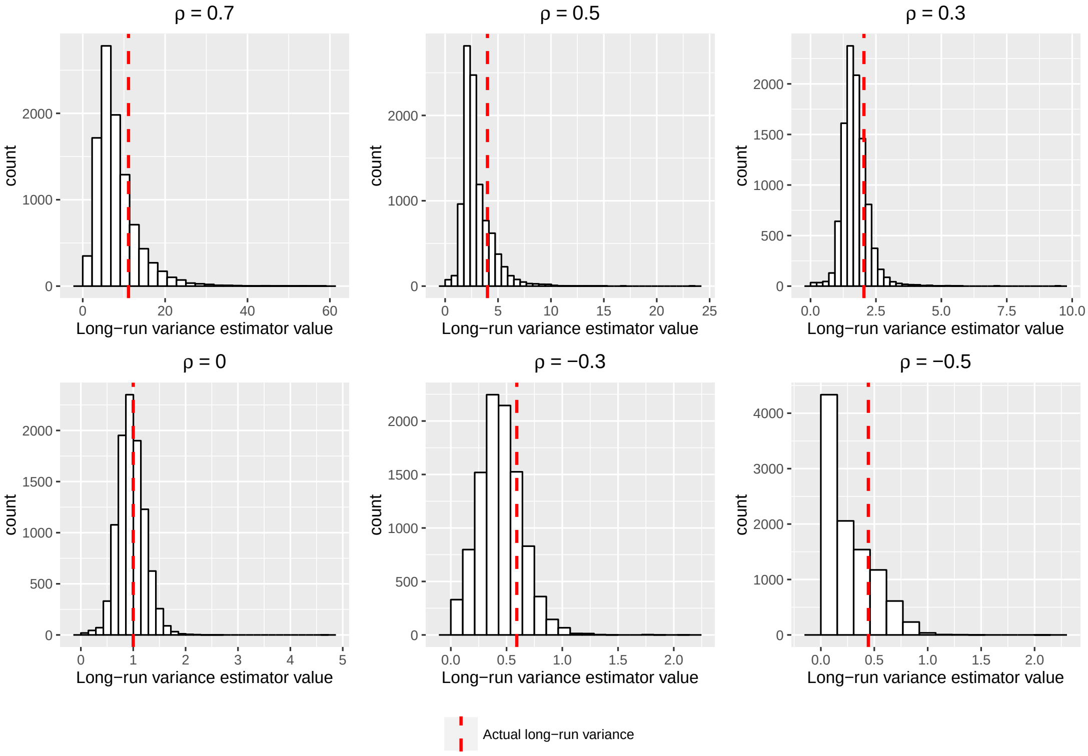
n = 100								
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean	Actual
0.7	0.06	4.92	6.97	10.20	17.92	56.91	8.20	11.11
0.5	0.06	2.08	2.57	3.45	5.53	23.36	2.95	4.00
0.3	0.06	1.40	1.65	1.94	2.49	9.53	1.70	2.04
0	0.06	0.79	0.95	1.12	1.40	4.72	0.96	1.00
-0.3	0.06	0.31	0.43	0.56	0.77	2.08	0.44	0.59
-0.5	0.06	0.06	0.20	0.42	0.70	2.00	0.27	0.44

n = 200								
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean	Actual
0.7	0.03	6.50	8.48	11.17	17.31	57.46	9.37	11.11
0.5	0.03	2.41	3.14	3.90	5.41	18.56	3.29	4.00
0.3	0.03	1.52	1.71	1.92	2.36	6.76	1.75	2.04
0	0.11	0.86	0.98	1.09	1.28	2.12	0.98	1.00
-0.3	0.03	0.36	0.45	0.54	0.70	1.16	0.45	0.59
-0.5	0.03	0.17	0.38	0.52	0.67	1.29	0.36	0.44

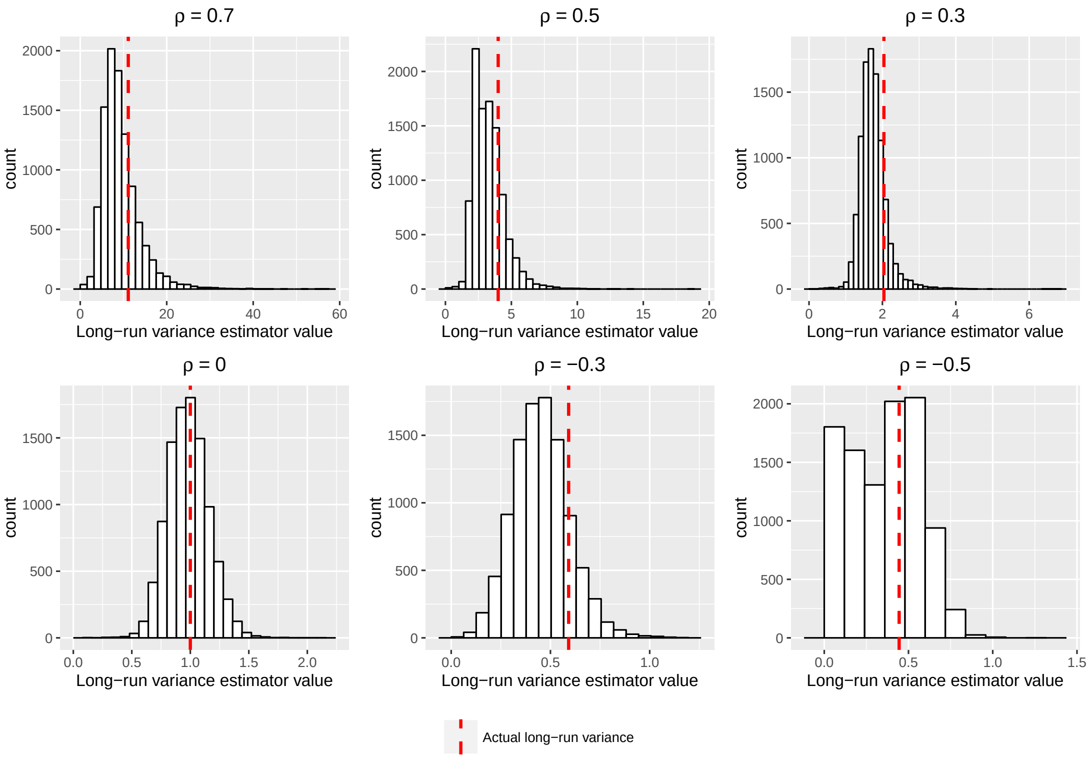
n = 500								
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean	Actual
0.7	1.84	8.13	9.67	11.67	15.48	51.61	10.17	11.11
0.5	0.58	3.11	3.50	4.04	5.02	12.49	3.62	4.00
0.3	0.65	1.62	1.75	1.90	2.36	4.39	1.80	2.04
0	0.45	0.92	0.99	1.07	1.18	1.67	0.99	1.00
-0.3	0.18	0.40	0.46	0.52	0.68	0.93	0.47	0.59
-0.5	0.01	0.39	0.46	0.53	0.61	0.85	0.45	0.44

Table 6.3: Tables illustrating the behavior and performance of our estimator $\hat{\sigma}^{+2}_{A,n}$ of σ^{*2}

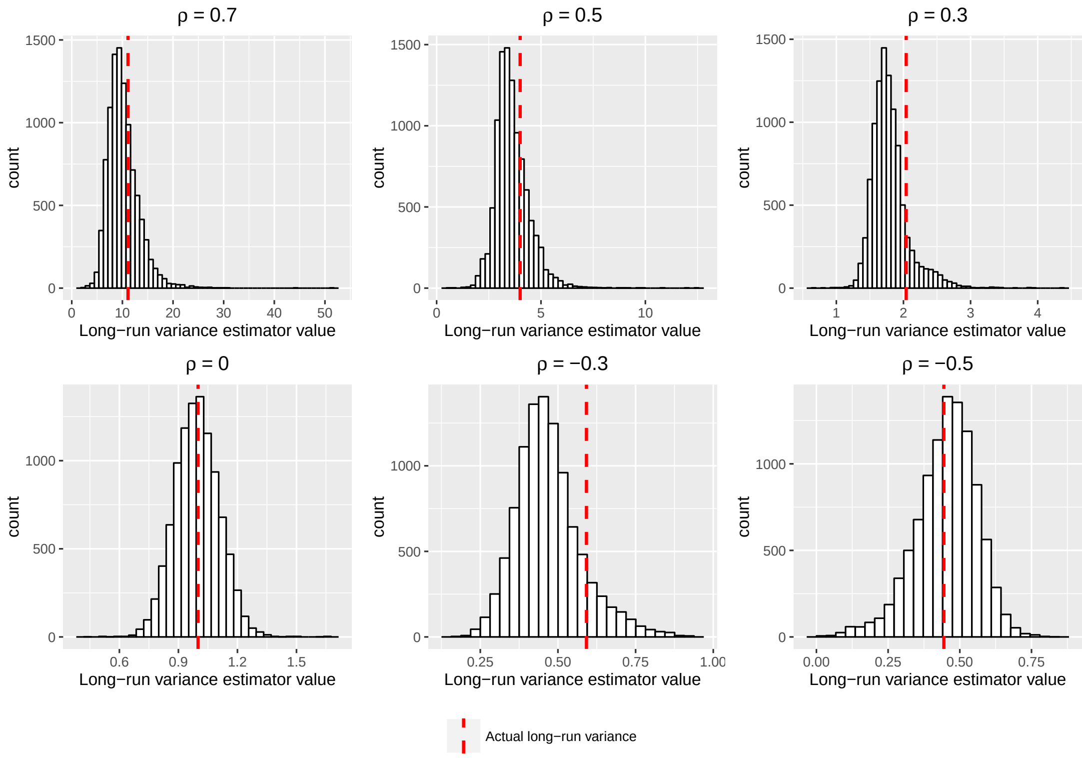
The performance of our long-run variance estimator can be examined even further using the histograms in Figure 6.2 that provide a graphical representation of its empirical distribution. Histogram divides the entire range of observed values into a series of intervals and counts how many of them fall into each interval. It is apparent that our long-run variance estimator $\hat{\sigma}^{+2}_{A,n}$ slightly underestimates σ^{*2}.


ρ = 0.7
ρ = 0.5
ρ = 0.3
ρ = 0
ρ = −0.3
ρ = −0.5
count
Long−run variance estimator value
Actual long−run variance

(a) n = 100


ρ = 0.7
ρ = 0.5
ρ = 0.3
ρ = 0
ρ = −0.3
ρ = −0.5
count
Long−run variance estimator value
Actual long−run variance

(b) n = 200



(a) n = 500

Figure 6.2: Histograms illustrating the performance of long-run variance estimator $\hat{\sigma}^{+2}_{A,n}$ for various n

6.1 Results of bootstrap procedures application

In this section we show the implementation and the results of our selected bootstrap procedures - block bootstrap with non-random block length (denoted as *circular bootstrap*), block bootstrap with random block length (denoted as *stationary bootstrap*) and *dependent wild bootstrap* (introduced in Shao (2010)). All of these methods were already discussed in the third chapter, now we will go into the details of their implementation. Again the simulations were performed using the **R** software.

6.1.1 Block bootstrap results

Parameter μ in model (4.1) was taken to be equal to zero, parameter ρ ranges in set $\{-0.5, -0.3, 0, 0.3, 0.5, 0.7\}$ and the sample lengths n range in set $\{100, 200, 500\}$.

For each combination of ρ and n the circular bootstrap simulation algorithm was set as follows:

1. A sample $y(1), \ldots, y(n)$ of length n is taken from model (4.1) under the null hypothesis, i.e. from $AR(1)$ model with parameter ρ.

2. The estimated optimal bandwidth $\hat{M}_A$ is then calculated from our sample according to the empirical rule-of-thumb described in section 4.3.1.

3. The estimate $\hat{\sigma}^{+2}_{A,n}$ based on trapezoidal flat-top kernel with bandwidth $\hat{M}_A$ is calculated as described in (4.17).

4. The estimated optimal block length $\hat{b}_{opt,CB}$ is calculated based on flat-top kernel estimators using the bandwidth $\hat{M}_A$ as described in (4.20).

5. The sample $y(1), \ldots, y(n)$ and block length $\hat{b}_{opt,CB}$ are used to create 1000 circular bootstrap samples as described for the (A) case of section 4.4. We obtain

$$y_1^*(1), \ldots, y_1^*(n)$$
$$\vdots$$
$$y_{1000}^*(1), \ldots, y_{1000}^*(n)$$

6. By calculation of the values of our test statistic T_n^S on the aforementioned bootstrap samples we obtain the bootstrapped values of our test statistic

$$T_{n,1}^{S*}, \ldots, T_{n,1000}^{S*}$$

where

$$T_{n,r}^{S*} = a(\log n) \cdot T_{n,r}^{A*} - b(\log n)$$

and

$$\begin{aligned} T_{n,r}^{A*} &= \max_{1 \le k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \cdot \frac{\left|S_{kn,r}^*\right|}{\hat{\sigma}_{A,n}^+} \right\}, \\ S_{kn,r}^* &= \sum_{t=1}^{k} (y_r^*(t) - \bar{y}_{n,r}^*), \\ \bar{y}_{n,r}^* &= \frac{1}{n} \sum_{l=1}^{n} y_r^*(l) \end{aligned}$$

7. We can now calculate the sample quantiles $C_{CB}^*(t), t \in (0,1)$ of the values $T_{n,1}^{S*}, \ldots, T_{n,1000}^{S*}$ and define a test of null hypothesis (5.1) based on circular bootstrap critical values on significance level $\alpha \in (0,1)$ as follows. Let T_n^S be the value of our test statistic on our original sample $y(1), \ldots, y(n)$. We reject H_0 if and only if

$$T_n^S > C_{CB}^*(1-\alpha). \tag{6.1}$$

The applied stationary bootstrap simulation algorithm was very similar, but using $\hat{b}_{opt,SB}$ as in (4.19) and (B) case of 4.4. Similarly we get sample quantiles $C^*_{SB}(t), t \in (0,1)$ and test the null hypothesis by comparison of

$$T_n^S > C^*_{SB}(1-\alpha). \tag{6.2}$$

The focus in the following section lies in the percentiles of the bootstrapped values of our test statistic obtained using circular bootstrap. For each combination of parameters ρ and sample length n we get 1000 sets of bootstrapped values of our test statistic, from which we can get 1000 observations of $C^*_{CB}(t)$ for any chosen $t \in (0,1)$. Following tables display the means of these observations for chosen values of t throughout the 1000 repetitions of circular bootstrap procedure. These can be compared to the ones in Table 6.1.

			n = 100		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.49	1.88	2.23	2.66	3.44
0.5	1.85	2.27	2.65	3.10	3.97
0.3	1.86	2.28	2.65	3.09	3.94
0	2.31	2.74	3.13	3.60	4.47
-0.3	4.37	5.02	5.63	6.23	7.15
-0.5	8.23	9.37	10.43	11.48	12.49

			n = 200		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.53	1.94	2.30	2.74	3.60
0.5	1.86	2.30	2.69	3.16	4.10
0.3	2.02	2.47	2.87	3.34	4.26
0	2.30	2.75	3.16	3.64	4.54
-0.3	4.11	4.75	5.34	6.05	6.96
-0.5	6.62	7.64	8.59	9.81	10.78

			n = 500		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.53	1.97	2.36	2.83	3.75
0.5	1.86	2.31	2.72	3.20	4.16
0.3	2.15	2.63	3.05	3.55	4.52
0	2.32	2.78	3.20	3.71	4.67
-0.3	3.80	4.40	4.97	5.65	6.77
-0.5	3.70	4.42	5.14	6.04	7.39

Table 6.4: Means of chosen percentiles of bootstrapped test statistics obtained using circular bootstrap

We can observe that with the exception of the i.i.d. case, i.e. when $\rho = 0$, the means of the percentiles obtained using circular bootstrap are lower than the ones in Table 6.1. We may also introduce the means of the percentiles obtained using stationary bootstrap procedure into the comparison. These are presented in Table 6.5 and are similar to the ones from circular bootstrap. The fact that the block bootstrap percentiles are in the mean lower than the quantiles of the simulations point out the possibility that the test procedures based on them might have trouble with keeping their prescribed level.

			n = 100		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.41	1.79	2.11	2.49	3.24
0.5	1.76	2.16	2.52	2.94	3.76
0.3	1.79	2.20	2.56	2.99	3.82
0	2.31	2.75	3.15	3.60	4.52
-0.3	4.49	5.17	5.80	6.40	7.49
-0.5	8.44	9.59	10.63	11.72	13.04

			n = 200		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.44	1.83	2.18	2.59	3.41
0.5	1.78	2.21	2.59	3.04	3.92
0.3	1.95	2.39	2.77	3.23	4.13
0	2.30	2.76	3.17	3.64	4.54
-0.3	4.24	4.89	5.49	6.20	7.23
-0.5	6.84	7.87	8.85	10.04	11.20

			n = 500		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.45	1.88	2.25	2.71	3.56
0.5	1.79	2.24	2.64	3.10	4.03
0.3	2.10	2.56	2.98	3.47	4.41
0	2.32	2.78	3.20	3.71	4.68
-0.3	3.90	4.52	5.09	5.79	7.00
-0.5	3.81	4.55	5.25	6.14	7.50

Table 6.5: Means of chosen percentiles of bootstrapped test statistics obtained using stationary bootstrap

Let us now take a look on our estimated optimal block length sizes $\hat{b}_{opt,CB}$ and $\hat{b}_{opt,SB}$. Table 6.6 describes the chosen block lengths in 1000 iterations of our block bootstrap procedures starting with the circular bootstrap.

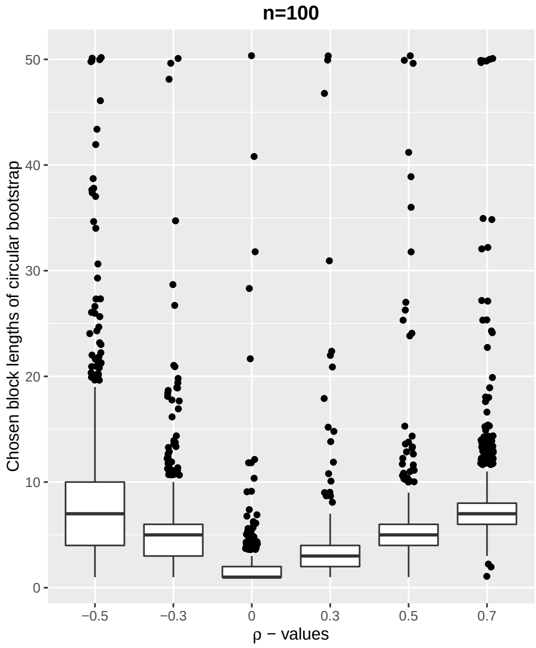
n = 100							
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	1.00	6.00	7.00	8.00	13.00	50.00	7.90
0.5	1.00	4.00	5.00	6.00	8.00	50.00	5.31
0.3	1.00	2.00	3.00	4.00	5.00	50.00	3.28
0	1.00	1.00	1.00	2.00	4.00	50.00	1.92
-0.3	1.00	3.00	5.00	6.00	10.00	50.00	5.11
-0.5	1.00	4.00	7.00	10.00	19.00	50.00	8.26

n = 200							
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	5.00	9.00	10.00	12.00	19.00	100.00	11.44
0.5	2.00	6.00	6.00	8.00	10.05	33.00	6.97
0.3	1.00	3.00	4.00	5.00	6.00	47.00	4.45
0	1.00	1.00	1.00	2.00	3.00	34.00	1.76
-0.3	1.00	3.00	5.00	7.00	11.00	78.00	5.66
-0.5	1.00	6.00	9.00	12.00	21.00	69.00	9.91

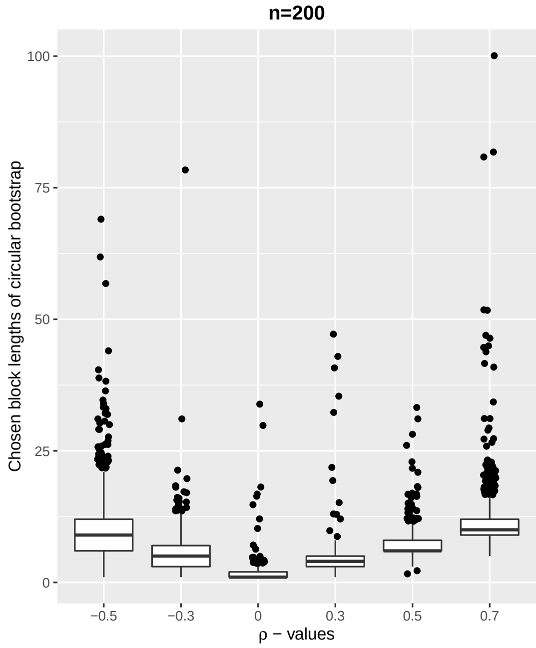
n = 500							
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	5.00	13.00	15.00	17.00	23.00	37.00	15.68
0.5	2.00	9.00	10.00	10.00	13.00	30.00	9.82
0.3	1.00	5.00	6.00	7.00	9.00	28.00	6.29
0	1.00	1.00	1.00	2.00	3.00	13.00	1.64
-0.3	1.00	5.00	7.00	9.00	11.00	19.00	6.83
-0.5	1.00	9.00	12.00	15.00	20.00	36.00	12.12

Table 6.6: Chosen block lengths for circular bootstrap

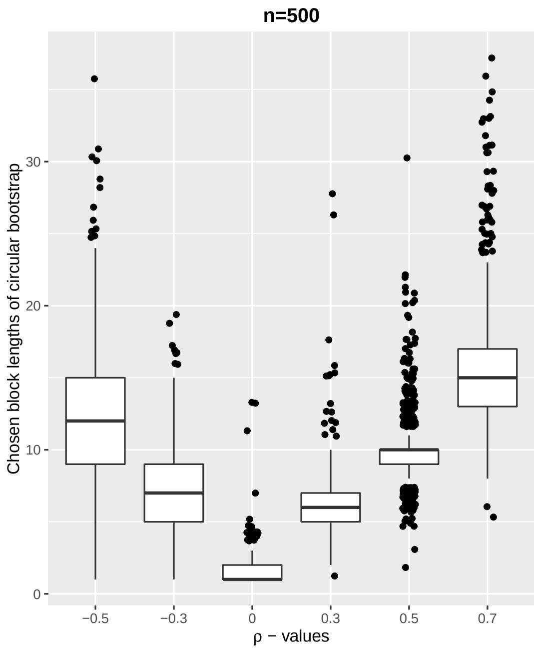
The chosen circular block lengths can be for different values of parameters compared also in boxplots displayed in Figure 6.3. We can observe that the maximum of our estimated optimal block lengths is often equal to $\frac{n}{2}$. This upper limit was set artificially to avoid cases when the chosen block length would be too high to create reasonable bootstrap samples. From the boxplots it is apparent that the cases when the artificial upper bound is used become less common with increasing sample size. It is also visible that estimated optimal block lengths increase with the absolute value of parameter ρ, i.e. we have higher block lengths for stronger dependence.



(a) n = 100



(b) n = 200



(c) n = 500

Figure 6.3: Boxplots illustrating the chosen block lengths for circular bootstrap with varying ρ and n

The estimated optimal block length sizes for circular bootstrap procedure appear in general lower than the ones for stationary bootstrap, which serve as the mean of the geometric distribution that the final block lengths are taken from.

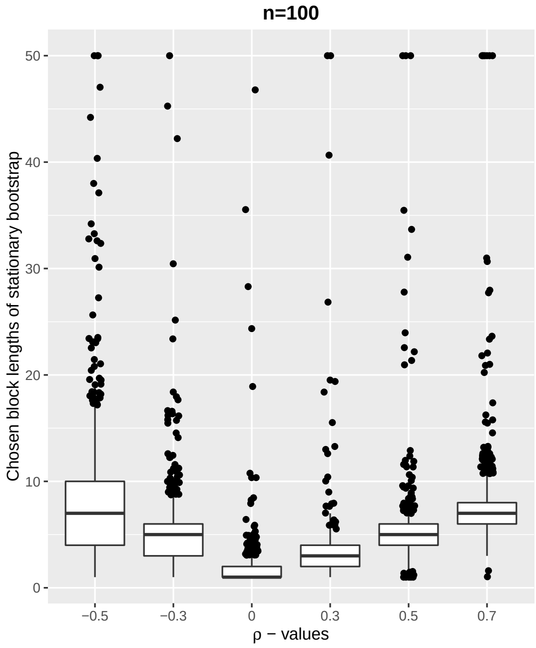
n = 100							
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	1.04	5.19	6.11	7.40	11.45	50.00	6.95
0.5	1.00	3.57	4.11	4.92	7.37	50.00	4.65
0.3	1.00	1.99	2.27	3.40	4.38	50.00	2.89
0	1.00	1.00	1.27	1.83	3.46	46.79	1.75
-0.3	1.00	2.80	3.95	5.13	8.99	50.00	4.48
-0.5	1.00	3.78	5.83	9.11	16.41	50.00	7.26

n = 200							
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	4.17	7.62	8.79	10.54	16.80	100.00	10.00
0.5	1.42	4.85	5.62	6.77	9.13	29.05	6.08
0.3	1.00	2.66	3.80	4.49	5.31	41.20	3.86
0	1.00	1.00	1.22	1.74	2.92	29.95	1.60
-0.3	1.00	3.04	4.52	6.33	9.49	68.54	4.95
-0.5	1.00	5.11	7.67	10.77	18.08	59.89	8.66

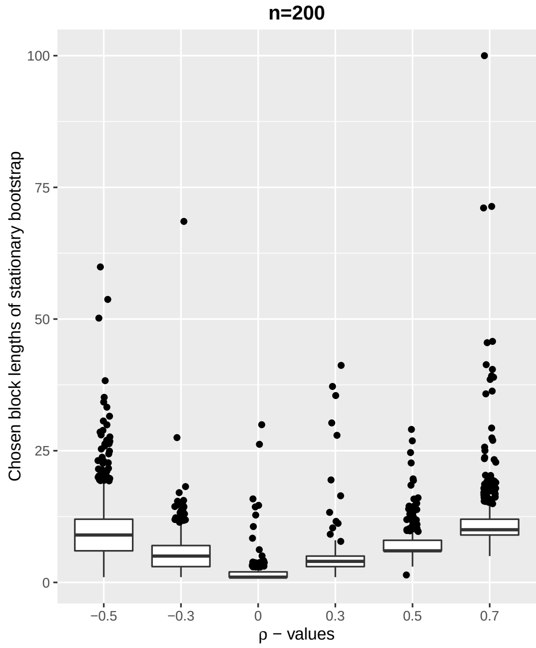
n = 500							
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	4.76	11.55	12.99	14.96	20.20	32.60	13.69
0.5	1.93	7.51	8.31	9.15	11.77	26.06	8.57
0.3	1.00	4.79	5.40	5.97	7.97	24.16	5.52
0	1.00	1.00	1.24	1.75	2.59	11.71	1.48
-0.3	1.00	4.23	5.88	7.59	10.04	16.97	5.97
-0.5	1.00	8.10	10.46	13.10	17.68	31.75	10.60

Table 6.7: Chosen block lengths for stationary bootstrap

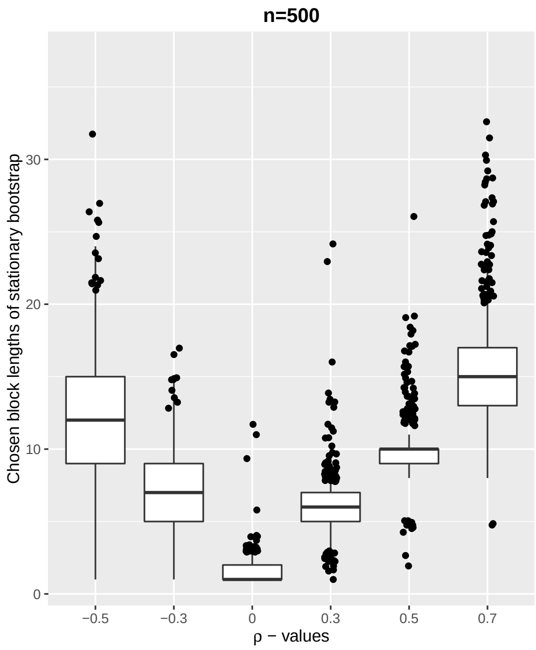
The chosen stationary bootstrap block lengths can be for different values of parameters compared also in boxplots displayed in Figure 6.4. Similarly to the circular bootstrap we can observe that the maximum of our estimated optimal block lengths is in a few cases equal to $\frac{n}{2}$. This upper limit was set artificially to avoid cases when the chosen block length would be too high to create reasonable bootstrap samples. From the boxplots it is apparent that the cases when the artificial upper bound is used become less common with increasing sample size. It is also visible that estimated optimal block lengths increase with the absolute value of parameter ρ, i.e. we have higher block lengths for stronger dependence.



(a) n = 100



(b) n = 200



(c) n = 500

Figure 6.4: Boxplots illustrating the chosen block lengths for stationary bootstrap with varying ρ and n

Dependent wild bootstrap implementation and results are the point of focus in the following section.

6.1.2 Dependent wild bootstrap results

According to Shao (2010) given the sample $y(1), \ldots, y(n)$ we construct the bootstrap time series as

$$y^*(t) = \bar{y}_n + (y(t) - \bar{y}_n) \cdot W_t,$$

where $\{W_t, t = 1, \dots, n\}$ are random variables with zero mean, unit variance and satisfy (3.2). These variables are used to model the underlying dependence structure of $\{y(t)\}_{t=1}^n$. There exists a variety of ways of choosing the model for $\{W_t, t = 1, \dots, n\}$, some of which may be found in Shao (2010). We decided to use the MA(m) model, i.e. variables $\{W_t, t = 1, \dots, n\}$ will mimic m-dependence structure. More specifically

$$W_t = b_0 \cdot \nu_t + b_1 \cdot \nu_{t-1} + \dots + b_m \cdot \nu_{t-m},$$

where $\{\nu_s\}_{-\infty}^{\infty}$ are i.i.d. normally distributed random variables with zero mean and unit variance. We decided to use uniform coefficients $b_0, \dots, b_m$, but they need to satisfy the assumptions of zero mean and unit variance of W_t. We now have

$$\mathsf{E}\ W_t = 0, \quad \operatorname{var}(W_t) = \sum_{j=0}^{m} b_j^2.$$

From here if we choose

$$b_j = \frac{\frac{1}{m+1}}{\sqrt{\sum\limits_{j=0}^{m+1} \frac{1}{(m+1)^2}}} = \frac{1}{\sqrt{m+1}}, \quad j = 0, \dots, m, \quad m > 0, \tag{6.3}$$

then W_t will have zero mean and unit variance. We still need to choose proper lag window m. In this case the autocovariance function of $\{W_t\}$ process is of the form

$$R_W(s) = \begin{cases} 1 - \frac{|s|}{m+1}, & |s| < m+1, \\ 0, & \text{otherwise.} \end{cases}$$

In Shao (2010) it is recommended that the autocovariance function of $\{W_t\}$ could be of the form $R_W(t) = \lambda\left(\frac{t}{l}\right)$, where l plays a similar role as block length in block bootstrap procedures. If we recall the Bartlett kernel $\lambda_B(\cdot)$ from (4.10) we can see that in our case

$$R_W(s) = \lambda_B\left(\frac{s}{m+1}\right).$$

Therefore we decided to make use of automatic bandwidth choice for the Bartlett kernel described in Andrews (1991) denoted $\hat{M}_{AND}$. This automatic approach is based on the approximation of dependence structure in the original sample by AR models. For the lag window m we will then use

$$\hat{m}_{AND} = \hat{M}_{AND} - 1.$$

The dependent wild bootstrap simulation algorithm was set as follows. Similarly to the previous cases again the parameter μ in the model (4.1) was set to be equal to zero, parameter ρ ranges in set $\{-0.5, -0.3, 0, 0.3, 0.5, 0.7\}$ and the sample lengths n range in set $\{100, 200, 500\}$. For each combination of ρ and n:

1. A sample $y(1), \ldots, y(n)$ of length n was taken from model (4.1) under the null hypothesis, i.e. from $AR(1)$ model.

2. An estimate of the optimal bandwidth $\hat{M}_A$ is then calculated from our sample according to the empirical rule-of-thumb described in section 4.3.1.

3. The estimate $\hat{\sigma}^{+2}_{A,n}$ based on trapezoidal flat-top kernel with bandwidth $\hat{M}_A$ is calculated as described in (4.17).

4. An estimate of the optimal bandwidth $\hat{M}_{AND}$ is then calculated from our sample according to the approach described in Andrews (1991) and used to calculate $\hat{m}_{AND}$.

5. We simulate 1000 samples $W_{1,r}, \ldots, W_{n,r}, r = 1, \ldots, 1000$ from MA($\hat{m}_{AND}$) process with coefficients as in (4.5), and use them to create our bootstrap observations

$$y^*_r(t) = \bar{y}_n + (y(t) - \bar{y}_n) \cdot W_{t,r}, \quad t = 1, \ldots, n, \quad r = 1, \ldots, 1000.$$

We obtain

$$\begin{gathered} y^*_1(1), \ldots, y^*_1(n) \\ \vdots \\ y^*_{1000}(1), \ldots, y^*_{1000}(n) \end{gathered}$$

6. By calculation of the values of our test statistic T^S_n on the aforementioned bootstrap samples we obtain the bootstrapped values of our test statistic

$$T^{S*}_{n,1}, \ldots, T^{S*}_{n,1000}.$$

7. Similarly to the block bootstrap procedures, we calculate the sample quantiles $C^*_{DWB}(t), t \in (0,1)$ of the values $T^{S*}_{n,1}, \ldots, T^{S*}_{n,1000}$ and define a test of our null hypothesis (5.1) based on dependent wild bootstrap critical values on significance level $\alpha \in (0,1)$ as follows. Let T^S_n be the value of our test statistic on our original sample $y(1), \ldots, y(n)$. We reject H_0 if and only if

$$T^S_n > C^*_{DWB}(1-\alpha). \tag{6.4}$$

The focus in the following section lies in the percentiles of the bootstrapped values of our test statistic obtained using dependent wild bootstrap. For each combination of parameters ρ and sample length n we get 1000 sets of bootstrapped values of our test statistic. From these we can get 1000 observations of $C^*_{DWB}(t)$ for any chosen $t \in (0,1)$. Table 6.8 displays the means of these observations for chosen values of t throughout the 1000 repetitions of circular bootstrap procedure. These can be compared to the ones presented in Table 6.1 and also to similar ones obtained using block bootstrap procedures displayed in Table 6.4 and Table 6.5.

			n = 100		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.80	2.41	2.97	3.64	4.94
0.5	2.25	2.88	3.46	4.17	5.55
0.3	2.43	3.08	3.68	4.41	5.82
0	2.92	3.62	4.27	5.07	6.58
-0.3	5.65	6.75	7.77	9.04	11.50
-0.5	10.82	12.70	14.41	16.49	20.51

			n = 200		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.78	2.40	2.95	3.65	4.98
0.5	2.26	2.91	3.52	4.25	5.69
0.3	2.54	3.21	3.84	4.61	6.12
0	2.90	3.61	4.28	5.10	6.67
-0.3	5.51	6.62	7.67	8.95	11.41
-0.5	9.46	11.19	12.78	14.71	18.42

			n = 500		
ρ	q(0.9)	q(0.95)	q(0.975)	q(0.99)	q(0.999)
0.7	1.79	2.42	3.00	3.71	5.07
0.5	2.27	2.94	3.56	4.32	5.82
0.3	2.72	3.42	4.07	4.89	6.48
0	2.91	3.61	4.27	5.09	6.67
-0.3	5.43	6.57	7.63	8.91	11.40
-0.5	5.80	7.03	8.18	9.58	12.28

Table 6.8: Means of chosen percentiles of bootstrapped test statistics obtained using dependent wild bootstrap

We can observe that the means of percentiles obtained using dependent wild bootstrap are much closer to the ones from Table 6.1 than it was in the case of block bootstrap procedures. In most cases they are even higher than the simulated ones. In this case it points out to the possibility that they might be even too prudent.

The bandwidth choices $\hat{M}_{AND}$ which occurred in our simulations are summarized in Table 6.9. Although the approach is different we can compare them to the automatic bandwidths $\hat{M}_A$ used for the flat-top kernels. The first one is based on approximations of the sample using AR processes and is designed for use with the Bartlett kernel, whereas the second one is based on the sample's correlogram and works well with flat-top kernels. We can observe that $\hat{M}_{AND}$ gives more stable results that are higher in the mean. Both of them behave similarly in terms of assigning broader lag windows in case of stronger dependency.

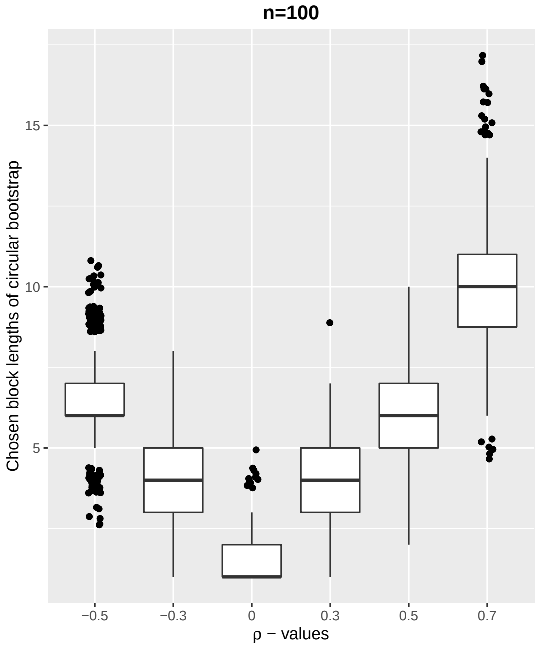
			n = 100				
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	5.00	8.75	10.00	11.00	13.00	17.00	9.78
0.5	2.00	5.00	6.00	7.00	8.00	10.00	6.16
0.3	1.00	3.00	4.00	5.00	6.00	9.00	3.82
0	1.00	1.00	1.00	2.00	3.00	5.00	1.61
-0.3	1.00	3.00	4.00	5.00	6.00	8.00	4.07
-0.5	3.00	6.00	6.00	7.00	9.00	11.00	6.47

			n = 200				
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	8.00	12.00	13.00	14.00	16.00	20.00	12.87
0.5	4.00	7.00	8.00	9.00	10.00	11.00	7.95
0.3	2.00	4.00	5.00	6.00	7.00	9.00	4.94
0	1.00	1.00	1.00	2.00	3.00	4.00	1.59
-0.3	1.00	5.00	5.00	6.00	7.00	8.00	5.11
-0.5	5.00	7.00	8.00	9.00	10.00	12.00	8.10

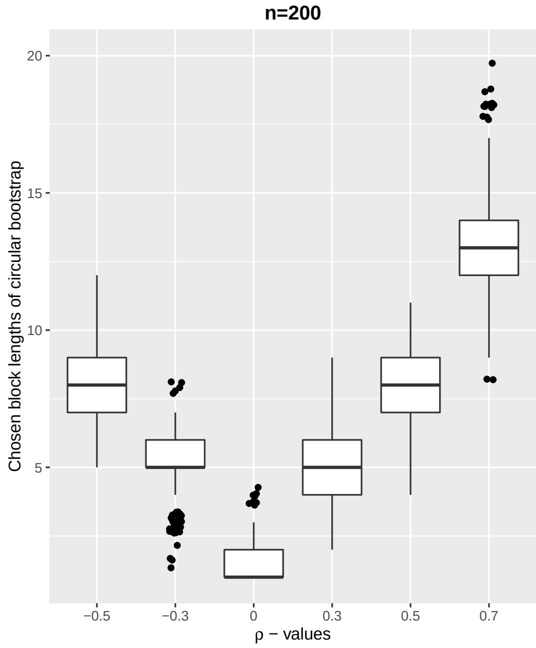
			n = 500				
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean
0.7	14.00	17.00	18.00	19.00	20.00	23.00	17.62
0.5	8.00	10.00	11.00	11.00	12.00	14.00	10.87
0.3	4.00	6.00	7.00	7.00	8.00	10.00	6.87
0	1.00	1.00	1.00	2.00	3.00	4.00	1.57
-0.3	5.00	6.00	7.00	7.00	8.00	9.00	6.88
-0.5	8.00	10.00	11.00	12.00	13.00	14.00	10.98

Table 6.9: Chosen bandwidths for dependent wild bootstrap

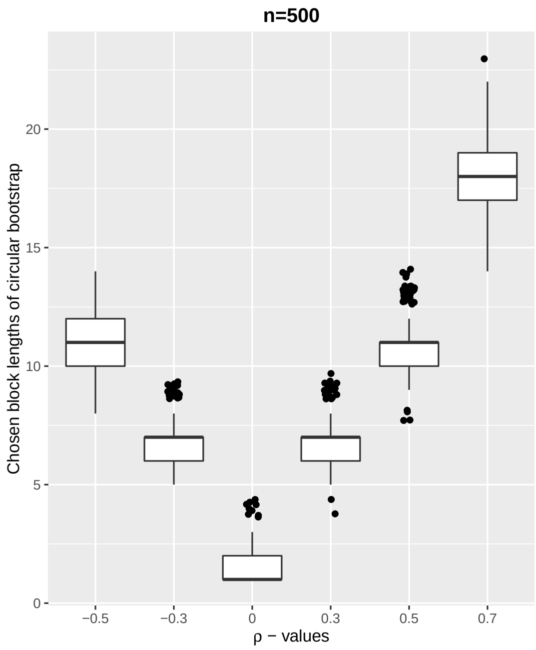
The choices of $\hat{M}_{AND}$ are illustrated in using boxplots displayed in 6.5. We can observe that when compared to $\hat{M}_A$ the number of observations beyond the whiskers of the boxplots is significantly lower, which implies higher stability.



(a) n = 100



(b) n = 200



(c) n = 500

Figure 6.5: Boxplots illustrating the chosen block lengths for circular bootstrap with varying ρ and n

Let us now take a look on the achieved levels of significance and power achieved in our simulations.

6.2 Simulated levels of significance

This section is devoted to illustrating the observed α-level of our discussed test procedures based on asymptotic, block bootstrap and dependent wild bootstrap critical values. Samples $y(1), \ldots, y(n)$ were generated under the null hypothesis and for each simulated sample the test procedures were performed. The relative numbers of false rejections of H_0 for each combination of parameters ρ and n are displayed. The discussed test procedures are denoted by the abbreviations - AS stands for test based on asymptotic critical values, CB for test based on circular bootstrap, SB for test based on stationary bootstrap and DWB for test based on dependent wild bootstrap critical values.

		n = 100			n = 200			n = 500		
			α			α			α	
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0.7	AS	0.018	0.007	0.006	0.008	0.006	0.005	0.018	0.003	0.001
	CB	0.244	0.115	0.044	0.189	0.083	0.040	0.163	0.092	0.047
	SB	0.268	0.130	0.049	0.202	0.094	0.042	0.166	0.092	0.047
	DWB	0.159	0.048	0.016	0.132	0.055	0.019	0.108	0.038	0.018
0.5	AS	0.030	0.005	0.005	0.032	0.008	0.002	0.025	0.003	0.001
	CB	0.193	0.102	0.042	0.172	0.083	0.047	0.153	0.076	0.042
	SB	0.218	0.113	0.047	0.183	0.090	0.045	0.155	0.080	0.042
	DWB	0.102	0.031	0.009	0.080	0.032	0.011	0.078	0.024	0.003
0.3	AS	0.028	0.003	0.001	0.036	0.011	0.004	0.052	0.020	0.005
	CB	0.166	0.088	0.052	0.146	0.075	0.038	0.152	0.080	0.046
	SB	0.177	0.091	0.048	0.163	0.082	0.033	0.157	0.082	0.045
	DWB	0.060	0.020	0.007	0.053	0.019	0.004	0.054	0.018	0.008
0	AS	0.039	0.005	0.002	0.038	0.012	0.005	0.036	0.012	0.003
	CB	0.115	0.058	0.024	0.112	0.053	0.019	0.101	0.058	0.024
	SB	0.112	0.053	0.021	0.108	0.053	0.022	0.101	0.054	0.024
	DWB	0.035	0.009	0.003	0.037	0.013	0.008	0.042	0.015	0.005
-0.3	AS	0.288	0.179	0.116	0.240	0.129	0.081	0.209	0.104	0.050
	CB	0.081	0.041	0.023	0.070	0.036	0.019	0.073	0.034	0.021
	SB	0.066	0.026	0.017	0.060	0.027	0.013	0.064	0.028	0.017
	DWB	0.005	0.001	0.001	0.005	0.000	0.000	0.006	0.002	0.001
-0.5	AS	0.610	0.536	0.461	0.429	0.354	0.291	0.182	0.096	0.058
	CB	0.069	0.040	0.024	0.117	0.064	0.040	0.076	0.048	0.025
	SB	0.055	0.028	0.017	0.098	0.052	0.029	0.066	0.044	0.022
	DWB	0.001	0.001	0.000	0.004	0.000	0.000	0.000	0.000	0.000

Table 6.10: Achieved α-level of discussed test procedures in the location model

We can see that the test based on asymptotic critical values (AS) manages to hold the level in cases when $\rho \geq 0$, but has trouble for $\rho \in \{-0.3, -0.5\}$ especially for lower n. This may be traced to the fact that our long-run variance estimator for these cases often hits the low artificial limit $\frac{2\pi}{n}$ which results in its underestimation. On the other hand all of the presented bootstrap procedures manage to hold the level for these values of ρ well. The block bootstrap procedures however have trouble with holding the level when $\rho \geq 0$ especially when ρ is high. Dependent wild bootstrap seems to hold the level better in these cases, however when $\rho < 0$ barely any rejections for the dependent wild bootstrap of the null hypothesis appear at all. This could imply lower power of this procedure in these cases of ρ when compared to block bootstrap procedures.

6.3 Simulated levels of power

Besides holding their prescribed levels of significance it is also important for the discussed test procedures to have high power, i.e. the probability that they reject the null hypothesis if it does not hold, should be close to 1. The following tables display the relative number of correct rejections of the null hypothesis under different alternatives - parameter δ that represents the change in the mean of our series varied in $\{0.5, 1.5\}$ and parameter q_n that represents the time point at which the change occurred varied in $\{\frac{n}{4}, \frac{n}{2}\}$. Similar abbreviations for the test procedures as in the previous section appear also in these tables.

		n = 100			n = 200			n = 500		
			α			α			α	
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0.7	AS	0.006	0.002	0.001	0.031	0.004	0.001	0.206	0.086	0.016
	CB	0.584	0.369	0.177	0.740	0.553	0.323	0.969	0.921	0.816
	SB	0.607	0.411	0.211	0.752	0.575	0.341	0.969	0.916	0.784
	DWB	0.511	0.281	0.120	0.683	0.481	0.221	0.951	0.850	0.663
0.5	AS	0.097	0.018	0.001	0.230	0.076	0.012	0.152	0.092	0.053
	CB	0.820	0.631	0.408	0.975	0.907	0.740	1.000	1.000	0.998
	SB	0.847	0.672	0.439	0.980	0.922	0.766	1.000	1.000	0.994
	DWB	0.719	0.418	0.136	0.950	0.796	0.519	1.000	0.998	0.977
0.3	AS	0.330	0.153	0.049	0.222	0.139	0.082	0.018	0.010	0.005
	CB	0.967	0.888	0.674	1.000	0.992	0.929	1.000	1.000	1.000
	SB	0.972	0.911	0.737	1.000	0.998	0.953	1.000	1.000	1.000
	DWB	0.890	0.644	0.271	0.995	0.969	0.881	1.000	1.000	1.000
0	AS	0.389	0.295	0.191	0.108	0.075	0.057	0.000	0.000	0.000
	CB	0.999	0.972	0.827	1.000	1.000	0.980	1.000	1.000	1.000
	SB	0.999	0.991	0.918	1.000	1.000	0.996	1.000	1.000	1.000
	DWB	0.985	0.862	0.612	1.000	0.999	0.987	1.000	1.000	1.000
-0.3	AS	0.398	0.323	0.181	0.074	0.046	0.032	0.000	0.000	0.000
	CB	1.000	0.998	0.873	1.000	1.000	0.996	1.000	1.000	1.000
	SB	1.000	1.000	0.976	1.000	1.000	1.000	1.000	1.000	1.000
	DWB	0.991	0.953	0.851	1.000	0.999	0.994	1.000	1.000	1.000
-0.5	AS	0.451	0.407	0.094	0.190	0.117	0.085	0.000	0.000	0.000
	CB	1.000	0.997	0.929	1.000	1.000	0.996	1.000	1.000	1.000
	SB	1.000	1.000	0.984	1.000	1.000	0.999	1.000	1.000	1.000
	DWB	0.988	0.925	0.796	1.000	0.998	0.990	1.000	1.000	1.000

Table 6.11: Achieved levels of power of our test procedures for $\delta = 1.5$ and $q_n = \frac{n}{4}$

		n = 100			n = 200			n = 500		
		α			α			α		
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0.7	AS	0.005	0.000	0.000	0.019	0.003	0.000	0.149	0.045	0.008
	CB	0.629	0.401	0.183	0.828	0.671	0.444	0.995	0.975	0.924
	SB	0.673	0.477	0.247	0.847	0.699	0.487	0.994	0.970	0.904
	DWB	0.538	0.271	0.098	0.772	0.564	0.354	0.979	0.929	0.852
0.5	AS	0.068	0.012	0.000	0.115	0.034	0.007	0.029	0.014	0.007
	CB	0.899	0.728	0.459	0.992	0.948	0.826	1.000	1.000	1.000
	SB	0.930	0.789	0.550	0.996	0.952	0.848	1.000	1.000	1.000
	DWB	0.824	0.607	0.360	0.981	0.931	0.825	1.000	0.998	0.994
0.3	AS	0.170	0.067	0.016	0.063	0.035	0.024	0.000	0.000	0.000
	CB	0.984	0.929	0.696	1.000	0.998	0.973	1.000	1.000	1.000
	SB	0.992	0.961	0.814	1.000	0.999	0.985	1.000	1.000	1.000
	DWB	0.968	0.883	0.700	0.998	0.993	0.975	1.000	1.000	0.999
0	AS	0.130	0.087	0.063	0.006	0.003	0.003	0.000	0.000	0.000
	CB	1.000	0.992	0.893	1.000	1.000	1.000	1.000	1.000	1.000
	SB	1.000	0.997	0.972	1.000	1.000	0.999	1.000	1.000	1.000
	DWB	0.999	0.989	0.944	1.000	1.000	1.000	1.000	1.000	1.000
-0.3	AS	0.098	0.076	0.030	0.005	0.003	0.002	0.000	0.000	0.000
	CB	1.000	1.000	0.943	1.000	1.000	1.000	1.000	1.000	1.000
	SB	1.000	1.000	0.993	1.000	1.000	1.000	1.000	1.000	1.000
	DWB	1.000	0.996	0.982	1.000	1.000	0.999	1.000	1.000	1.000
-0.5	AS	0.134	0.123	0.020	0.007	0.006	0.001	0.000	0.000	0.000
	CB	1.000	0.999	0.970	1.000	1.000	1.000	1.000	1.000	1.000
	SB	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	DWB	0.999	0.985	0.960	1.000	1.000	0.998	1.000	1.000	1.000

Table 6.12: Achieved levels of power of our test procedures for $\delta = 1.5$ and $q_n = \frac{n}{2}$

		n = 100			n = 200			n = 500		
		α			α			α		
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0.7	AS	0.017	0.010	0.009	0.014	0.003	0.000	0.047	0.007	0.001
	CB	0.289	0.147	0.059	0.323	0.183	0.094	0.418	0.256	0.151
	SB	0.315	0.164	0.068	0.334	0.185	0.091	0.427	0.256	0.154
	DWB	0.195	0.065	0.020	0.225	0.115	0.044	0.354	0.212	0.113
0.5	AS	0.061	0.018	0.001	0.118	0.034	0.009	0.380	0.187	0.071
	CB	0.344	0.207	0.120	0.454	0.303	0.187	0.701	0.591	0.467
	SB	0.372	0.222	0.129	0.479	0.309	0.188	0.707	0.586	0.445
	DWB	0.228	0.106	0.041	0.354	0.182	0.100	0.611	0.461	0.317
0.3	AS	0.126	0.031	0.005	0.350	0.178	0.060	0.837	0.670	0.474
	CB	0.437	0.288	0.173	0.623	0.484	0.369	0.948	0.897	0.835
	SB	0.446	0.304	0.178	0.623	0.489	0.364	0.951	0.888	0.822
	DWB	0.281	0.141	0.066	0.484	0.317	0.186	0.867	0.755	0.623

		n = 100			n = 200			n = 500		
			α			α			α	
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0	AS	0.262	0.085	0.012	0.648	0.404	0.220	0.987	0.970	0.923
	CB	0.559	0.419	0.294	0.865	0.767	0.647	0.998	0.997	0.990
	SB	0.556	0.413	0.293	0.865	0.770	0.648	0.998	0.997	0.989
	DWB	0.382	0.245	0.150	0.691	0.537	0.376	0.990	0.960	0.923
-0.3	AS	0.667	0.480	0.274	0.927	0.844	0.697	1.000	0.997	0.990
	CB	0.581	0.412	0.288	0.886	0.808	0.728	1.000	1.000	1.000
	SB	0.537	0.364	0.237	0.868	0.785	0.679	1.000	1.000	1.000
	DWB	0.354	0.211	0.110	0.729	0.555	0.392	0.990	0.961	0.909
-0.5	AS	0.633	0.497	0.441	0.906	0.772	0.547	0.998	0.995	0.981
	CB	0.468	0.319	0.201	0.930	0.850	0.754	1.000	1.000	1.000
	SB	0.415	0.254	0.149	0.916	0.814	0.679	1.000	1.000	1.000
	DWB	0.270	0.121	0.045	0.703	0.494	0.341	0.988	0.951	0.893

Table 6.13: Achieved levels of power of our test procedures for $\delta = 0.5$ and $q_n = \frac{n}{4}$

		n = 100			n = 200			n = 500		
			α			α			α	
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0.7	AS	0.022	0.015	0.010	0.012	0.004	0.003	0.040	0.009	0.000
	CB	0.265	0.146	0.061	0.285	0.159	0.065	0.335	0.212	0.134
	SB	0.296	0.170	0.070	0.292	0.163	0.064	0.344	0.209	0.121
	DWB	0.197	0.066	0.027	0.196	0.086	0.036	0.255	0.139	0.074
0.5	AS	0.052	0.008	0.004	0.085	0.023	0.006	0.277	0.120	0.033
	CB	0.302	0.165	0.078	0.360	0.219	0.139	0.586	0.446	0.340
	SB	0.318	0.179	0.085	0.377	0.226	0.129	0.590	0.444	0.311
	DWB	0.168	0.061	0.016	0.263	0.109	0.037	0.474	0.292	0.170
0.3	AS	0.105	0.018	0.004	0.259	0.111	0.038	0.706	0.531	0.356
	CB	0.359	0.236	0.129	0.523	0.379	0.261	0.847	0.769	0.679
	SB	0.395	0.238	0.128	0.528	0.384	0.250	0.852	0.759	0.665
	DWB	0.182	0.078	0.027	0.351	0.171	0.088	0.730	0.561	0.411
0	AS	0.189	0.054	0.012	0.519	0.331	0.138	0.963	0.903	0.802
	CB	0.447	0.312	0.198	0.733	0.621	0.511	0.989	0.980	0.963
	SB	0.449	0.311	0.207	0.732	0.609	0.507	0.990	0.979	0.958
	DWB	0.261	0.140	0.062	0.531	0.359	0.249	0.948	0.880	0.782
-0.3	AS	0.584	0.380	0.224	0.896	0.773	0.617	0.999	0.996	0.985
	CB	0.442	0.284	0.173	0.806	0.672	0.545	1.000	1.000	0.995
	SB	0.402	0.238	0.143	0.784	0.641	0.493	1.000	0.999	0.993
	DWB	0.203	0.088	0.035	0.534	0.320	0.175	0.967	0.894	0.808
-0.5	AS	0.668	0.542	0.470	0.840	0.657	0.450	1.000	0.996	0.972
	CB	0.348	0.198	0.123	0.796	0.665	0.517	1.000	0.998	0.987
	SB	0.274	0.149	0.082	0.743	0.590	0.419	1.000	0.998	0.986
	DWB	0.115	0.031	0.005	0.452	0.231	0.104	0.929	0.836	0.695

Table 6.14: Achieved levels of power of our test procedures for $\delta = 0.5$ and $q_n = \frac{n}{2}$

We can observe that the best results in terms of achieved power are given by block bootstrap methods, the dependent wild bootstrap gives slightly worse results and the worst ones come from the test procedure based on asymptotic critical values. In general the best results in terms of power are achieved for lower values of ρ, which can be traced to the fact that with the increase of ρ, the actual long-run variance σ^{*2} also increases. There is also another interesting point worth mentioning. For the case when $\delta = 1.5$ and $\rho \in \{0, -0.3, -0.5\}$ the achieved power of the asymptotic test procedure decreases with the sample size. For the case of very large sample lengths $n = 500$ the null hypothesis is not rejected once by this procedure, even though it does not hold. This can be traced to our long-run variance estimator $\hat{\sigma}^{+2}_{A,n}$ which in the situation of a large change in the mean significantly overestimates long-run variance of our error sequence as is demonstrated in Table 6.15.

			n = 100					
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean	Actual
0.7	1.12	7.64	11.93	18.84	37.04	87.75	14.98	11.11
0.5	1.35	4.28	6.53	10.48	21.19	46.43	8.46	4.00
0.3	1.18	2.85	5.16	9.42	17.23	29.07	6.85	2.04
0	1.13	2.38	5.02	8.38	13.53	21.64	5.91	1.00
-0.3	0.84	2.67	5.15	7.90	11.70	18.47	5.57	0.59
-0.5	0.46	2.74	4.41	6.95	10.38	17.92	5.02	0.44

			n = 200					
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean	Actual
0.7	0.15	11.18	16.72	25.94	52.98	120.37	21.28	11.11
0.5	2.41	6.58	11.19	19.55	36.94	60.59	14.61	4.00
0.3	1.46	7.54	13.62	21.04	29.47	45.72	14.72	2.04
0	1.53	10.69	15.64	20.50	26.56	37.29	15.55	1.00
-0.3	0.98	11.18	15.09	18.42	23.81	30.02	14.77	0.59
-0.5	2.12	8.66	13.19	17.21	21.34	28.73	12.98	0.44

			n = 500					
ρ	Min	q(0.25)	Median	q(0.75)	q(0.95)	Max	Mean	Actual
0.7	7.09	16.98	26.72	47.67	90.48	160.27	36.05	11.11
0.5	4.15	23.31	37.19	53.27	72.78	121.30	38.91	4.00
0.3	4.60	34.93	43.48	53.98	66.41	85.04	43.78	2.04
0	17.41	40.27	45.38	50.91	59.83	74.34	45.76	1.00
-0.3	24.70	41.69	46.00	50.42	56.76	70.93	46.04	0.59
-0.5	20.85	40.38	44.33	48.87	54.54	63.15	44.35	0.44

Table 6.15: Tables illustrating the behavior of our long-run variance estimator in the case of $\delta = 1.5$ and $q_n = \frac{n}{4}$

We can see that the estimator is not able to handle a large change in the mean and gives unreasonably high values even for large sample sizes. This significantly decreases the values of our test statistic to the point that they are lower than the asymptotic critical values in all 1000 repetitions of our procedure when $\rho \in \{0, -0.3, -0.5\}$. On the other hand, the bootstrap procedures manage to cope with this problem well. This is caused by the fact that in bootstrap the estimator of the long-run variance not only affects the test statistic, but the critical values as well. Therefore they are able to reflect it. As an effort to improve the behavior of the estimator under the alternative hypothesis, we also tried to use an analogy of a more robust estimator $\tilde{\sigma}_n^{*2}(L)$ discussed in Antoch et al. (1997). However it significantly underestimated long-run variance under the null hypothesis.

Chapter 7

Simulation study in linear model

To examine the behavior of bootstrap procedures in more complex models we prepared a simulation study on linear regression model with a change in parameters and error terms that form a linear process.

7.1 Model definition

We are interested in the following model

$$y(t) = \boldsymbol{x}(t)^\top \boldsymbol{\beta}_n + \boldsymbol{x}(t)^\top \boldsymbol{\delta}_n \cdot I\{t > q_n\} + e(t), \quad t = 1, \dots, n, \tag{7.1}$$

where $q_n(\leq n)$, $\boldsymbol{\beta}_n = (\beta_{1,n}, \dots, \beta_{p,n})^\top$ and $\boldsymbol{\delta}_n = (\delta_{1,n}, \dots, \delta_{p,n})^\top \neq \mathbf{0}$ are unknown parameters, $\boldsymbol{x}(t) = (\boldsymbol{x}_1(i), \dots, \boldsymbol{x}_p(t))^\top, x_1(t) = 1, i = 1, \dots, n$ are known design points. We will assume that the error terms $e(t)$ satisfy condition [b.1.] from the fourth chapter.

7.2 Null hypothesis and test statistic

The interest lies in testing the hypothesis

$$H_0 : q_n = n \text{ against } H_1 : q_n < n.$$

Again the null hypothesis corresponds to the situation where the regression parameters did not change in the observed sample. Similarly to the first chapter our test statistic will be based on the partial sums

$$S^*_{kn} = \sum_{t=1}^{k} \left(y(t) - \boldsymbol{x}(t)^\top \hat{\beta}_n \right), \quad k = 1, \dots, n,$$

where

$$\hat{\beta}_n = \boldsymbol{C}_{nn}^{-1} \sum_{t=1}^{n} \boldsymbol{x}(t) y(t), \quad C_{kn} = \sum_{t=1}^{k} \boldsymbol{x}(t) \boldsymbol{x}(\boldsymbol{t})^\top, \quad k = 1, \dots, n,$$
$$C^0_{kn} = C_{kn} - C_{nn}, \quad k = 1, \dots, n.$$

The statistic is in fact calculated from the least squares residuals

$$\hat{e}(t) = y(t) - \boldsymbol{x}(t)^\top \hat{\boldsymbol{\beta}}_n, \quad t = 1, \ldots, n.$$

In order to reflect the variability of the residuals, we have to standardize these partial sums by an estimator of residual long-run variance. The idea behind the estimator (4.17) introduced for the location model may be utilized also in this case. However, we have to make one crucial adjustment - the calculation will be based on our L_2-residuals rather than on the observations of our dependent variable. Let us denote $\hat{R}_e(k), k = -n+1, \ldots, n-1$ the estimated autocovariance function of our calculated residuals similarly as in (4.6). Let

$$\hat{\sigma}^2_{e,n}(M) = \sum_{k=-n+1}^{n-1} \lambda_T \left(\frac{k}{M}\right) \hat{R}_e(k)$$

and $\hat{M}_{e,A}$ be the automatic bandwidth choice obtained using the approach described in 4.3.1 on $\hat{R}_e(k)$. Furthermore, let

$$\hat{\sigma}^{+2}_{e,A,n} = \max\left\{\frac{2\pi}{n}, \hat{\sigma}^2_{e,n}(\hat{M}_{e,A})\right\}.$$

and

$$T^A_{e,n} = \max_{1 \le k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \cdot \frac{|S_{e,kn}|}{\hat{\sigma}^+_{e,A,n}} \right\}. \tag{7.2}$$

Finally, the standardized version of this test statistic will be denoted as

$$T^S_{e,n} = a(\log n) \cdot T^A_{e,n} - b_1(\log n). \tag{7.3}$$

7.3 Bootstrap procedures

When compared to location model, we will have to make adjustments also to the applied bootstrap procedures. The main idea remains the same - sampling blocks of our data for block procedures and using artificial random variables to distort our data from the fitted values in the case of dependent wild bootstrap procedure.

7.3.1 Block bootstrap methods

In the case of location model we were creating blocks from observations $y(1), \ldots, y(n)$. Now we have the data in the form

$$\begin{pmatrix} y(1) \\ \boldsymbol{x}(1) \end{pmatrix}, \ldots, \begin{pmatrix} y(n) \\ \boldsymbol{x}(n) \end{pmatrix}$$

and in the simulations we will use a modification of pair bootstrap, i.e. we will form bootstrap samples consisting of blocks

$$\begin{pmatrix} y(t_m) \\ \boldsymbol{x}(t_m) \end{pmatrix}, \ldots, \begin{pmatrix} y(t_m + b_m - 1) \\ \boldsymbol{x}(t_m + b_m - 1) \end{pmatrix}, \quad m = 1, \ldots, L,$$

where L stands for the number of blocks, t_m the first observation of m-th block and b_m the length of m-th block. Similarly to the case of the location model, we will use an estimate of the optimal block length based on flat-top kernel. However, now it will be calculated from the residual series $\hat{e}(i)$ and will be denoted $\hat{b}_{e,CB}$ for circular and $\hat{b}_{e,SB}$ for stationary bootstrap.

7.3.2 Dependent wild bootstrap

We will utilize the idea behind the dependent wild bootstrap used in the location model also in the linear model. However we will again have to make a few crucial adjustments. Given the observations

$$\begin{pmatrix} y(1) \\ \boldsymbol{x}(1) \end{pmatrix}, \ldots, \begin{pmatrix} y(n) \\ \boldsymbol{x}(n) \end{pmatrix}$$

we will first calculate the fitted values

$$\hat{y}(t) = \boldsymbol{x}(t)^\top \hat{\boldsymbol{\beta}}_n, \quad t = 1, \ldots, n.$$

and define dependent wild bootstrap pseudoseries as

$$\begin{pmatrix} y^*(1) \\ \boldsymbol{x}(1) \end{pmatrix}, \ldots, \begin{pmatrix} y^*(n) \\ \boldsymbol{x}(n) \end{pmatrix},$$

where

$$y^*(t) = y(t) + (y(t) - \hat{y}(t)) \cdot W_t, \quad t = 1, \ldots, n.$$

Again $\{W_t, i = 1, \ldots, n\}$ are random variables with zero mean, unit variance and satisfy (3.2). In the simulation study these variables are taken from MA(m) model with uniform coefficients, where the optimal m is estimated using the bandwidth estimator $\hat{M}_{e,AND}$ introduced in Andrews (1991) as it was in the case of the location model with one difference - $\hat{M}_{e,AND}$ is calculated from the residuals $\hat{e}(t)$. The final estimate of optimal choice for m is then given by $\hat{m}_{e,AND} = \hat{M}_{e,AND} - 1$.

7.4 Simulation setup

We chose to perform a simulation study on the following linear model with a change in parameters

$$y(t) = 1 + x(t) + \delta_1 \cdot I\{t > q_n\} + \delta_2 \cdot x(t) \cdot I\{t > q_n\} + e(t), \quad t = 1, \ldots, n. \quad (7.4)$$

Regressors $x(t), t = 1, \ldots, n$ were taken from uniform $U(0, 2)$ distribution and error terms $e(t), t = 1, \ldots, n$ were taken from autoregressive AR(1) model with zero mean univariate innovations and parameter $\rho \in (-1, 1)$. The focus will lie on the behavior and performance of our test procedures - AS based on critical values 5.1, CB based on circular bootstrap, SB based on stationary bootstrap and DWB based on dependent wild bootstrap. In our simulations parameter ρ ranges in the same set as in the location model $\{-0.5, -0.3, 0, 0.3, 0.5, 0.7\}$ and sample length n ranges in $\{100, 200, 500\}$.

7.5 Simulated levels of significance

In the following table we present the observed α-level of our test procedures. We generated 1000 samples

$$\begin{pmatrix} y(1) \\ \boldsymbol{x}(1) \end{pmatrix}, \ldots, \begin{pmatrix} y(n) \\ \boldsymbol{x}(n) \end{pmatrix}$$

under the null hypothesis and applied all four test procedures on each. The relative number of false rejections of H_0 for each combination of ρ and n is displayed in Table 7.1 with the aforementioned notation.

		n = 100			n = 200			n = 500		
			α			α			α	
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0.7	AS	0.028	0.019	0.013	0.013	0.004	0.001	0.015	0.002	0.000
	CB	0.232	0.118	0.043	0.207	0.121	0.062	0.176	0.090	0.048
	SB	0.248	0.132	0.052	0.219	0.120	0.052	0.189	0.099	0.050
	DWB	0.075	0.019	0.004	0.069	0.026	0.009	0.049	0.012	0.004
0.5	AS	0.034	0.016	0.008	0.030	0.011	0.002	0.022	0.002	0.000
	CB	0.209	0.113	0.048	0.159	0.087	0.041	0.134	0.067	0.035
	SB	0.224	0.116	0.054	0.165	0.086	0.047	0.140	0.063	0.034
	DWB	0.067	0.023	0.007	0.052	0.024	0.011	0.060	0.019	0.007
0.3	AS	0.032	0.011	0.003	0.030	0.007	0.002	0.044	0.011	0.003
	CB	0.179	0.096	0.052	0.146	0.071	0.037	0.139	0.080	0.035
	SB	0.185	0.100	0.054	0.154	0.074	0.039	0.145	0.078	0.036
	DWB	0.057	0.017	0.004	0.061	0.017	0.003	0.063	0.024	0.013
0	AS	0.037	0.012	0.004	0.036	0.009	0.003	0.042	0.007	0.002
	CB	0.110	0.052	0.024	0.115	0.058	0.027	0.102	0.060	0.023
	SB	0.098	0.052	0.025	0.116	0.051	0.024	0.103	0.049	0.024
	DWB	0.040	0.014	0.004	0.056	0.028	0.010	0.045	0.023	0.011
-0.3	AS	0.277	0.173	0.116	0.239	0.131	0.063	0.192	0.102	0.045
	CB	0.079	0.044	0.020	0.068	0.028	0.015	0.082	0.040	0.020
	SB	0.064	0.027	0.012	0.050	0.019	0.011	0.066	0.032	0.016
	DWB	0.007	0.001	0.000	0.004	0.000	0.000	0.014	0.003	0.002
-0.5	AS	0.603	0.518	0.453	0.388	0.314	0.263	0.199	0.127	0.076
	CB	0.086	0.039	0.018	0.076	0.043	0.022	0.098	0.055	0.033
	SB	0.070	0.025	0.015	0.061	0.035	0.015	0.088	0.046	0.031
	DWB	0.000	0.000	0.000	0.006	0.001	0.000	0.005	0.000	0.000

Table 7.1: Achieved α-level of discussed test procedures in linear model with a change in parameters

We can observe that the results are very similar to the case of the location model. Again the test based on asymptotic critical values manages to hold the level in cases when $\rho \geq 0$, but has trouble with $\rho \in \{-0.3, -0.5\}$ especially for lower n. This has a similar cause as in the simulation study in the location model. For lower values of n the estimator $\hat{\sigma}^{+2}_{e,A,n}$ hits the lower artificial bound and underestimates our residual long-run variance. Our test statistic then oversteps the asymptotic critical values, which as opposed to bootstrap critical values are not able to adapt. Block bootstrap methods do not manage to hold their level for higher values of ρ, but perform well for the other cases. Dependent wild bootstrap on the other hand seems even too prudent especially for lower values of ρ.

7.6 Simulated levels of power

Besides holding their prescribed level a well designed test procedure should be also able to reliably detect when the null hypothesis does not hold. The following tables display the relative number of correct rejections of the null hypothesis under different alternatives - due to the amount of time required for a set of simulations to complete we examined the following two alternatives $\boldsymbol{\delta} = (0.5, 0)^\top$, i.e. an increase in intercept by 0.5, and $\boldsymbol{\delta} = (0, 0.5)^\top$, i.e. an increase in slope by 0.5. Change point varied in $q_n \in \{\frac{n}{4}, \frac{n}{2}\}$.

		n = 100			n = 200			n = 500		
			α			α			α	
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0.7	AS	0.018	0.010	0.007	0.011	0.003	0.002	0.048	0.014	0.002
	CB	0.298	0.157	0.074	0.271	0.157	0.077	0.319	0.225	0.126
	SB	0.323	0.174	0.072	0.282	0.168	0.084	0.336	0.226	0.131
	DWB	0.080	0.018	0.004	0.112	0.029	0.006	0.171	0.081	0.037
0.5	AS	0.057	0.021	0.010	0.091	0.023	0.002	0.285	0.126	0.041
	CB	0.306	0.190	0.088	0.379	0.242	0.145	0.582	0.456	0.345
	SB	0.330	0.205	0.102	0.399	0.246	0.134	0.582	0.457	0.326
	DWB	0.116	0.035	0.013	0.180	0.083	0.039	0.396	0.236	0.134
0.3	AS	0.107	0.021	0.006	0.236	0.112	0.035	0.694	0.512	0.351
	CB	0.365	0.233	0.143	0.521	0.371	0.254	0.845	0.761	0.674
	SB	0.373	0.238	0.156	0.527	0.389	0.253	0.847	0.760	0.655
	DWB	0.173	0.067	0.025	0.330	0.175	0.084	0.691	0.527	0.390
0	AS	0.198	0.056	0.010	0.496	0.309	0.153	0.950	0.896	0.771
	CB	0.458	0.327	0.225	0.729	0.605	0.490	0.983	0.965	0.936
	SB	0.457	0.327	0.220	0.735	0.602	0.483	0.985	0.968	0.935
	DWB	0.302	0.180	0.088	0.574	0.409	0.281	0.935	0.860	0.778
-0.3	AS	0.606	0.404	0.235	0.880	0.759	0.597	0.998	0.993	0.979
	CB	0.448	0.300	0.177	0.811	0.703	0.561	0.998	0.996	0.986
	SB	0.398	0.248	0.145	0.791	0.649	0.492	0.998	0.994	0.986
	DWB	0.225	0.111	0.056	0.563	0.368	0.223	0.951	0.878	0.787
-0.5	AS	0.642	0.525	0.467	0.844	0.658	0.441	0.999	0.992	0.976
	CB	0.335	0.207	0.118	0.807	0.651	0.501	0.999	0.999	0.992
	SB	0.285	0.157	0.075	0.748	0.578	0.413	0.999	0.999	0.990
	DWB	0.146	0.042	0.011	0.471	0.271	0.123	0.952	0.865	0.725

Table 7.2: Achieved levels of power for the alternative $\delta = (0.5, 0)^\top$ and $q_n = \frac{n}{4}$

		n = 100			n = 200			n = 500		
			α			α			α	
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0.7	AS	0.018	0.011	0.010	0.017	0.001	0.001	0.055	0.016	0.001
	CB	0.279	0.159	0.085	0.310	0.186	0.093	0.400	0.279	0.172
	SB	0.304	0.175	0.097	0.334	0.196	0.090	0.418	0.280	0.166
	DWB	0.124	0.042	0.010	0.148	0.079	0.024	0.255	0.141	0.074
0.5	AS	0.071	0.017	0.007	0.105	0.024	0.003	0.383	0.189	0.072
	CB	0.351	0.225	0.124	0.440	0.307	0.184	0.728	0.601	0.452
	SB	0.387	0.239	0.132	0.452	0.310	0.181	0.736	0.584	0.429
	DWB	0.195	0.081	0.035	0.289	0.159	0.082	0.570	0.400	0.277
0.3	AS	0.126	0.030	0.006	0.336	0.161	0.058	0.816	0.664	0.490
	CB	0.433	0.294	0.188	0.632	0.508	0.364	0.944	0.890	0.816
	SB	0.450	0.309	0.199	0.648	0.510	0.348	0.945	0.881	0.796
	DWB	0.268	0.147	0.079	0.476	0.311	0.200	0.848	0.718	0.574
0	AS	0.248	0.092	0.022	0.680	0.468	0.249	0.988	0.969	0.922
	CB	0.541	0.396	0.304	0.861	0.778	0.705	0.998	0.994	0.990
	SB	0.549	0.400	0.291	0.853	0.776	0.692	0.997	0.995	0.989
	DWB	0.418	0.271	0.167	0.745	0.612	0.486	0.988	0.970	0.914
-0.3	AS	0.666	0.458	0.262	0.942	0.857	0.727	1.000	0.994	0.983
	CB	0.589	0.420	0.280	0.940	0.849	0.752	1.000	0.999	0.998
	SB	0.539	0.374	0.243	0.921	0.813	0.717	1.000	0.999	0.998
	DWB	0.396	0.233	0.141	0.794	0.637	0.474	0.993	0.968	0.935
-0.5	AS	0.658	0.527	0.464	0.911	0.776	0.567	0.999	0.995	0.982
	CB	0.498	0.344	0.218	0.920	0.839	0.744	1.000	1.000	0.999
	SB	0.442	0.277	0.158	0.907	0.805	0.686	1.000	1.000	1.000
	DWB	0.328	0.173	0.077	0.761	0.571	0.412	0.986	0.950	0.891

Table 7.3: Achieved levels of power for the alternative $\delta = (0.5, 0)^\top$ and $q_n = \frac{n}{2}$

		n = 100			n = 200			n = 500		
			α			α			α	
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0.7	AS	0.021	0.013	0.012	0.018	0.006	0.002	0.053	0.014	0.002
	CB	0.304	0.150	0.071	0.304	0.173	0.090	0.392	0.255	0.154
	SB	0.317	0.168	0.081	0.317	0.183	0.089	0.405	0.256	0.140
	DWB	0.101	0.021	0.004	0.128	0.044	0.009	0.197	0.088	0.035
0.5	AS	0.050	0.016	0.006	0.094	0.026	0.009	0.276	0.120	0.047
	CB	0.319	0.173	0.097	0.384	0.236	0.141	0.619	0.487	0.345
	SB	0.346	0.192	0.090	0.396	0.255	0.147	0.624	0.473	0.333
	DWB	0.129	0.042	0.010	0.198	0.075	0.031	0.420	0.248	0.125
0.3	AS	0.095	0.019	0.004	0.234	0.105	0.036	0.695	0.518	0.332
	CB	0.383	0.251	0.151	0.514	0.363	0.244	0.836	0.757	0.660
	SB	0.392	0.266	0.151	0.522	0.367	0.246	0.839	0.765	0.648
	DWB	0.180	0.080	0.028	0.320	0.175	0.089	0.688	0.510	0.364

		n = 100			n = 200			n = 500		
			α			α			α	
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0	AS	0.180	0.055	0.009	0.502	0.294	0.140	0.958	0.894	0.773
	CB	0.435	0.314	0.205	0.732	0.613	0.485	0.984	0.967	0.947
	SB	0.437	0.303	0.202	0.730	0.607	0.491	0.988	0.970	0.941
	DWB	0.280	0.141	0.074	0.573	0.400	0.280	0.945	0.874	0.778
-0.3	AS	0.571	0.374	0.210	0.890	0.739	0.582	1.000	0.995	0.982
	CB	0.425	0.290	0.183	0.798	0.657	0.514	1.000	0.996	0.994
	SB	0.390	0.253	0.151	0.767	0.617	0.469	1.000	0.997	0.990
	DWB	0.223	0.109	0.060	0.562	0.357	0.192	0.962	0.885	0.792
-0.5	AS	0.650	0.529	0.460	0.842	0.645	0.441	1.000	0.998	0.988
	CB	0.353	0.212	0.126	0.792	0.644	0.512	1.000	0.994	0.987
	SB	0.290	0.166	0.080	0.725	0.581	0.436	1.000	0.996	0.987
	DWB	0.148	0.044	0.012	0.469	0.272	0.129	0.938	0.849	0.712

Table 7.4: Achieved levels of power for the alternative $\delta = (0, 0.5)^\top$ and $q_n = \frac{n}{4}$

		n = 100			n = 200			n = 500		
			α			α			α	
ρ	method	0.1	0.05	0.025	0.1	0.05	0.025	0.1	0.05	0.025
0.7	AS	0.021	0.013	0.011	0.010	0.003	0.001	0.057	0.015	0.002
	CB	0.306	0.176	0.071	0.295	0.178	0.090	0.400	0.272	0.163
	SB	0.332	0.196	0.081	0.320	0.182	0.086	0.403	0.266	0.149
	DWB	0.126	0.045	0.008	0.150	0.066	0.022	0.248	0.134	0.072
0.5	AS	0.047	0.013	0.005	0.114	0.022	0.005	0.391	0.198	0.070
	CB	0.351	0.208	0.111	0.441	0.277	0.175	0.723	0.587	0.482
	SB	0.371	0.227	0.108	0.456	0.286	0.169	0.727	0.589	0.463
	DWB	0.179	0.084	0.032	0.283	0.149	0.068	0.585	0.433	0.300
0.3	AS	0.092	0.018	0.003	0.309	0.147	0.056	0.817	0.659	0.480
	CB	0.414	0.285	0.166	0.588	0.458	0.333	0.931	0.872	0.799
	SB	0.437	0.292	0.165	0.597	0.454	0.338	0.926	0.869	0.791
	DWB	0.283	0.159	0.072	0.439	0.297	0.186	0.830	0.711	0.573
0	AS	0.261	0.097	0.020	0.654	0.425	0.238	0.987	0.966	0.909
	CB	0.552	0.411	0.291	0.832	0.742	0.648	0.997	0.993	0.987
	SB	0.555	0.409	0.291	0.836	0.748	0.653	0.996	0.994	0.988
	DWB	0.430	0.280	0.171	0.736	0.582	0.446	0.982	0.950	0.908
-0.3	AS	0.650	0.433	0.250	0.946	0.865	0.731	1.000	0.999	0.989
	CB	0.561	0.390	0.255	0.912	0.828	0.736	1.000	1.000	1.000
	SB	0.522	0.346	0.217	0.898	0.805	0.705	1.000	1.000	1.000
	DWB	0.395	0.242	0.130	0.773	0.631	0.486	0.991	0.958	0.910
-0.5	AS	0.656	0.520	0.441	0.907	0.744	0.524	0.998	0.994	0.979
	CB	0.511	0.324	0.198	0.919	0.846	0.733	1.000	1.000	1.000
	SB	0.441	0.261	0.150	0.899	0.789	0.658	1.000	1.000	1.000
	DWB	0.338	0.179	0.084	0.708	0.540	0.376	0.990	0.957	0.886

Table 7.5: Achieved levels of power for the alternative $\delta = (0, 0.5)^\top$ and $q_n = \frac{n}{2}$

We can see that the results are similar in all examined cases of alternatives. Similarly to the location model under the $\delta = 0.5$ alternative, for higher values of

ρ block bootstrap procedures provide the best results in terms of achieved power. However, none of the procedures are able to reliably reject the null hypothesis even for sample size $n = 500$. For lower values of ρ all of the procedures provide satisfactory results.

Conclusion

The aim of this thesis was to explore the possible uses of bootstrap procedures in the change point detection problem and compare their performance to the asymptotic approach. In the first chapters we provided the reader with the necessary introduction into off-line and on-line setting of the discussed problem and a short review of bootstrap procedures designed specifically to handle dependent data. Location model was then used to illustrate the practical issues one has to tackle when implementing kernel based estimators and bootstrap methods, i.e. the choice of kernel, bandwidth, block lengths and other parameters. Simulated study in this model provided us with an insight into the performance of our estimator of long-run variance, automatic parameter choices and most importantly asymptotic and bootstrap test procedures. The last chapter was aimed on the the extension of the discussed approaches used in location model to the case of a more complex linear model. Simulation study was performed on a typical regression line model with a possible change in intercept and slope. Both of these simulation studies show that bootstrap procedures are superior to the asymptotic approach both in terms of holding their level and their power. These advantages however come at the cost of computational simplicity.

The issues we had to tackle came typically from the choice of parameters of our automatic choices or our calculation methods. For example several ways of choosing proper parameter m for our MA(m) model in dependent wild bootstrap were considered. Our final choice was then affected also by the similarity of the autocovariance function of our process to the Bartlett kernel. Also a more robust method of estimating sample autocovariance function that was designed to handle possible change in parameters was considered, but did not provide better results and therefore was not finally used for the sake of computational simplicity.

Utilizing bootstrap procedures in the setting of change point detection problem is a very broad topic. Naturally, the next point of focus could be the application of the discussed procedures also in the on-line setting, we could also examine the effect different methods of long-run variance estimation have on our results, other forms of dependence in the residuals is also worth examining and the comparison between automatic parameter choices and fixed ones might prove useful too.

Bibliography

Donald WK Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, pages 817–858, 1991.

J Antoch and M Hušková. Detection of structural changes in regression. *Tatra Mt. Math. Publ*, 26:201–215, 2003.

J Antoch, M Hušková, and Z Prášková. Effect of dependence on statistics for determination of change. *Journal of Statistical Planning and Inference*, 60(2): 291–310, 1997.

Peter J Brockwell and Richard A Davis. *Time series: theory and methods.* Springer Science & Business Media, 2013.

Miklós Csörgö and Lajos Horváth. *Limit theorems in change-point analysis*, volume 18. John Wiley & Sons Inc, 1997.

Lajos Horváth, Marie Hušková, Piotr Kokoszka, and Josef Steinebach. Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, 126 (1):225–251, 2004.

Marie Hušková and Claudia Kirch. Bootstrapping sequential change-point tests for linear regression. *Metrika*, 75(5):673–708, 2012.

Claudia Kirch. Block permutation principles for the change analysis of dependent data. *Journal of Statistical Planning and Inference*, 137(7):2453–2474, 2007.

Jens-Peter Kreiss and Efstathios Paparoditis. Bootstrap methods for dependent data: A review. *Journal of the Korean Statistical Society*, 40(4):357–378, 2011.

Soumendra N Lahiri. Theoretical comparisons of block bootstrap methods. *Annals of Statistics*, pages 386–404, 1999.

Dimitris N Politis. On nonparametric function estimation with infinite-order flat-top kernels. *Probability and Statistical Models with applications*, pages 469–483, 2001.

Dimitris N Politis. Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, 15(4-5):517–533, 2003.

Dimitris N Politis. Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory*, 27(04):703–744, 2011.

Dimitris N Politis and Joseph P Romano. A circular block-resampling procedure for stationary data. *Exploring the limits of bootstrap*, pages 263–270, 1992.

Dimitris N Politis and Joseph P Romano. The stationary bootstrap. *Journal of the American Statistical association*, 89(428):1303–1313, 1994.

Dimitris N Politis and Halbert White. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

Joseph P Romano and Lori A Thombs. Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association*, 91(434):590–600, 1996.

Murray Rosenblatt. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43, 1956.

Xiaofeng Shao. The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235, 2010.

Chien-Fu Jeff Wu. Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics*, pages 1261–1295, 1986.

List of Figures

List of Tables